Nonparametric Partial Identification of Causal Net and Mechanism Average Treatment Effects*

Carlos A. Flores† Alfonso Flores-Lagunes‡

July 9, 2010

Abstract

When analyzing the causal effect of a treatment on an outcome it is important to understand the mechanisms or channels through which the treatment works. In this paper we study net and mechanism average treatment effects ($NATE$ and $MATE$, respectively), which provide an intuitive decomposition of the total average treatment effect (ATE) that enables learning about how the treatment affects the outcome. We derive informative nonparametric bounds for these two effects allowing for heterogeneous effects and without requiring the use of an instrumental variable or having an outcome with bounded support. We employ assumptions requiring weak monotonicity of mean potential outcomes within or across subpopulations defined by the potential values of the mechanism variable under each treatment arm. We illustrate the identifying power of our bounds by analyzing what part of the ATE of a training program on weekly earnings and employment is due to the obtainment of a GED, high school, or vocational degree.

Key words and phrases: causal inference, treatment effects, net effects, direct effects, nonparametric bounds, principal stratification.

JEL classification: C13, C21, C14

*Helpful comments have been provided by Ken Chay, Fabrizia Mealli, seminar participants at the University of Arizona, Carleton University, CEPS/INSTEAD, and Georgia State University, and conference participants at the 2008 Midwest Econometrics Group, the 2008 Latin American Meetings of the Econometric Society, the 2009 New York Camp Econometrics, the 2009 European Meetings of the Econometric Society, the 2009 American Economic Association Annual Meeting, and the 2009 IRP Summer Workshop. NSF funding under grants SES-0852211 and SES-0852139 is gratefully acknowledged. The first author also acknowledges financial support through the James W. McLamore Summer Research Awards in Business and the Social Sciences from the University of Miami. Competent research assistance was provided by Maria Bampasidou. All errors are our own.

†Department of Economics, University of Miami. Email: caflores@miami.edu

‡Food and Resource Economics Department and Economics Department, University of Florida, and IZA, Bonn, Germany. Email: alfonsofl@ufl.edu

1 Introduction

An important topic in econometrics is the estimation of the average effect of a treatment or intervention on an outcome. When analyzing this effect, it is also important to understand the mechanisms or channels through which the treatment affects the outcome.[1] In this paper we study net and mechanism average treatment effects ($NATE$ and $MATE$, respectively), which provide an intuitive decomposition of the average treatment effect (ATE) that enables learning about how the treatment causally affects the outcome. We derive informative nonparametric bounds for these two effects in a heterogeneous effects setting without requiring the use of an instrumental variable or having an outcome with bounded support. Our approach is based on three sets of assumptions. The first assumes that the treatment is randomly assigned and imposes an individual-level monotonicity assumption of the effect of the treatment on the variable representing the mechanism. The other two sets of assumptions place inequality restrictions on the mean potential outcomes of specific subpopulations defined by the potential values of the mechanism variable. One set imposes those restrictions within subpopulations, while the second imposes them across subpopulations. Importantly, the specific assumptions in these two sets can be combined, changed, and some even dropped, depending on their plausibility, identifying power, and the economic theory behind any given application.

Identification of net and mechanism effects is a difficult task since it requires stronger conditions than those necessary to identify total treatment effects (e.g., Robins and Greenland, 1992; Rubin, 2004; Petersen et al., 2006). Intuitively, the mechanism variable would be endogenous in a regression of the outcome on the treatment and the mechanism variable. Even if the treatment is randomly assigned, there is non-random selection into the different values of the mechanism variable, so individuals with different values of the mechanism are not comparable, and a comparison of their potential outcomes does not yield a causal effect. The assumptions currently available in the literature to point identify net average treatment effects involve strong unconfoundedness assumptions requiring the mechanism to be "exogenous" conditional on covariates, plus other functional form, distributional, or constant treatment effects assumptions (e.g., Robins and Greenland, 1992; Petersen et al., 2006; Flores and Flores-Lagunes, 2009; Imai et al., 2010). In this paper, we follow the alternative strategy of deriving bounds for these causal effects under weaker assumptions than those required for point identification.

We derive our results within the Principal Stratification (PS) framework introduced by Frangakis and Rubin (2002), which has its roots in the analysis of identification of causal effects using instrumental variables in Imbens and Angrist (1994) and Angrist, Imbens and Rubin (1996). PS provides a framework for studying causal treatment effects when controlling

[1] Examples of empirical papers concerned with learning about the importance of a given mechanism (or controlling for it when estimating average treatment effects) include Angrist and Chen (2008), Black and Smith (2004), Currie and Moretti (2003), Dearden et al. (2002), and Simonsen and Skipper (2006).

for a variable that has been affected by the treatment—in our case the mechanism variable. The basic idea behind PS is to compare treated and control individuals in the same "principal strata", meaning that they share the same potential values of the post-treatment variable. Since the strata an individual belongs to is not affected by the treatment assignment, the comparison of potential outcomes within strata yields a causal effect.

Following previous literature on partial identification of net effects (Kaufman et al., 2005; Cai et al., 2008; Sjölander, 2009), we assume that the treatment is randomly assigned and that both the treatment and the mechanism variable of interest are binary. Concentrating on this canonical case allows us to focus on the main ideas behind our partial identification results and to set the basis for extensions to other settings. Moreover, this is an important case in practice. Most of the program evaluation literature focus on the binary-treatment case (e.g., Imbens and Wooldridge, 2009), and binary mechanism variables are relevant in practice, as is the case in our empirical application. Additionally, randomized experiments have gained importance in many fields in economics as a way of estimating average causal effects, such as in labor (e.g., Heckman et al., 1999) and development economics (e.g., Duflo et al., 2008). In this context, the methods we develop can be employed to analyze the role of potential causal mechanisms of the treatment under study, as illustrated in the empirical application of section 4.

We start our analysis in the following section by defining our parameters of interest. The $NATE$ equals the average potential outcome from a counterfactual treatment in which the effect of the original treatment on the mechanism variable of interest is blocked, minus the average potential outcome under the control treatment. The $MATE$ equals the difference between the ATE and the $NATE$. We show that, regardless of the treatment assignment, the typical data contains information on the first potential outcome used in the definition of $NATE$ only for a particular subpopulation: those individuals for which the treatment does not affect the mechanism variable. We derive nonparametric bounds for the $NATE$ of this subpopulation by assuming that the treatment is randomly assigned and by imposing an individual-level monotonicity assumption on the effect of the treatment on the mechanism variable, which is a condition also imposed by existing methods for estimation of net effects. These two assumptions have been previously used to derive bounds for net effects (Kaufman et al., 2005; Cai et al., 2008; Sjölander, 2009), and are also common in other settings (Imbens and Angrist, 1994; Zhang and Rubin, 2003; Zhang et al., 2008; Lee, 2009).

Our key insight in the derivation of bounds for the population $NATE$ and $MATE$ is to write them as a function of mean potential outcomes in each of the strata defined by the potential values of the mechanism variable under each treatment arm. Then, we relate the (partially or point) identified mean potential outcomes of the different strata in the population to those that are unidentified. To this end, we present two additional sets of assumptions involving weak inequalities of mean potential outcomes for specific strata. The first set of assumptions

imposes weak mean inequality restrictions for the different potential outcomes within a given strata. An example of these assumptions is that the $MATE$ for the subpopulation whose mechanism variable is affected by the treatment is non-negative. Assumptions involving weak monotonicity of individual-level potential outcomes have been used previously to bound (total) treatment effects (Manski, 1997). The assumptions we propose here are weaker than similar assumptions previously used in the literature to bound net effects (Sjölander, 2009) by requiring weak monotonicity to hold at the strata rather than at the individual level. Our bounds based on this set of assumptions are sharper than those currently available in the statistics literature, and they do not restrict the outcome to have a bounded support.

The second set of assumptions we consider imposes weak mean inequality restrictions for a given potential outcome across strata. These assumptions have not been considered before to derive bounds for $NATE$ and $MATE$ (to our knowledge), and we show they can have substantial identifying power. An example of these assumptions is that the mean potential outcomes for the subpopulation whose mechanism is affected by the treatment is always less than or equal to the corresponding average potential outcomes for the subpopulation who always has a high value of the mechanism variable regardless of the treatment assigned. Assumptions involving weak monotonicity of mean potential outcomes across specific subpopulations have been considered in other settings (Manski and Pepper, 2000; Zhang and Rubin, 2003; Zhang et al., 2008). We derive nonparametric bounds for the population $NATE$ and $MATE$ under each of these two additional sets of assumptions separately, and also combining them.

Most of the recent work on net or direct effects, which we briefly review in the next section, has been outside the field of economics. Our work, however, is related to two recent papers in the economics literature. Lee (2009) and Zhang et al. (2008) derive bounds for the effect of a randomly-assigned training program on wages considering the fact that wages are only observed for those individuals who are employed. It relates to the present paper since employment status may be regarded as a mechanism through which training affects wages. Both papers derive nonparametric bounds for the average treatment effect (ATE) of training on wages for the subpopulation of individuals who would be employed whether they received training or not. To derive bounds for the $NATE$ of the subpopulation for which the mechanism variable is not affected by the treatment, we use the same strategy as Zhang et al. (2008), which is similar in spirit to that in Lee (2009). Our paper differs from Lee (2009) and Zhang et al. (2008) in important ways. First, the set up is different, since in those papers the observability of the outcome (wages) depends on an intermediate variable (employment status), while in ours the outcome is always observed. Second, the question and hence the parameters of interest are different. Those papers focus on the ATE of a training program on wages while controlling for selection into employment, whereas our focus is on $NATE$ and $MATE$ in order to decompose

the ATE and study how the treatment affects the outcome.[2] Another key difference is that we derive nonparametric bounds for the population $NATE$ and $MATE$, and not only for the local average net effect of the subpopulation they focus on.

Two other contributions of the paper are as follows. Recently, the PS approach as applied to the study of net (or direct) effects (Mealli and Rubin, 2003; Rubin, 2004, 2005) has been criticized in the statistics literature because of its focus on estimating the net effect only for those individuals whose mechanism variable is not affected by the treatment (Robins et al., 2007; Joffe et al., 2007; VanderWeele, 2008). In this context, a contribution of this paper is to show how the PS approach can be employed to derive bounds for the population $NATE$ and $MATE$.

The second contribution relates to the current debate between "reduce form" and "structural" models. In the last two decades there has been a "credibility revolution" in empirical economics based on the so-called causal literature, which emphasizes the identification of causal effects and pays careful attention to the internal validity of the estimators and the study design (Imbens, 2010; Angrist and Pischke, 2010). This literature is sometimes criticized in favor of structural models on the grounds of being reduced form and thus not allowing a deeper understanding of the causal process (mechanism) behind the estimated causal effects (Deaton, 2010a,b; Heckman, 2010; Heckman and Urzua, 2010; Keane, 2010).[3] Here, we go beyond the reduced-form effect and provide a way to analyze the channels through which the treatment affects the outcome within the causal literature framework. We view our work as a step towards bridging those two views.

The rest of the paper is organized as follows. Section 2 presents the parameters of interest and briefly reviews the related statistics literature. Section 3 presents the main identification results of the paper. In Section 4 we illustrate the identifying power of the bounds derived in the paper by analyzing what part of the average treatment effect of the Job Corps training program on weekly earnings and employment is due to the obtainment of a high school, GED, or vocational degree. Our results suggest that obtaining such a degree accounts for at most fifty (sixty) percent of the total average effect of the program on employment (weekly earnings). Section 5 concludes. The proofs of our main results are presented in the appendix.

2 Definition of Parameters and Literature Review

Assume we have a random sample of size n from a large population. For each unit i in the sample, let $T_i \in \{0, 1\}$ indicate whether the unit received the treatment of interest ($T_i = 1$)

[2] Nevertheless, as discussed later, in the subpopulation for which the treatment does not affect the mechanism the average treatment and the average net effects are equal.

[3] A common critique is an "excessive" focus on the total effect of the treatment on an outcome, usually ignoring thinking about "how and why things work" (Deaton, 2010a).

or the control treatment ($T_i = 0$). We analyze the part of the effect of T on an outcome Y that works through a mechanism variable S. Since S is affected by the treatment, we denote by $S_i(\tau)$ for $\tau = 0, 1$ the potential values of the mechanism variable. Hence, $S_i(1)$ and $S_i(0)$ represent the value of the mechanism variable individual i would receive if exposed to treatment or not, respectively. At this stage, we do not restrict S to be binary.

Define the "composite" potential outcomes $Y_i(\tau, \zeta)$, where the first argument refers to one of the treatment arms ($\tau \in \{0, 1\}$) and the second argument represents one of the potential values of the mechanism variable S ($\zeta \in \{S_i(0), S_i(1)\}$). Note that the potential outcomes $Y_i(1, S_i(1))$ and $Y_i(0, S_i(0))$ correspond to the potential outcomes $Y_i(1)$ and $Y_i(0)$ typically used in the literature to define treatment effects. The potential outcome $Y_i(1, S_i(0))$ represents the outcome individual i would receive if she were exposed to the treatment but the effect of the treatment on the mechanism were blocked by keeping the mechanism at $S_i(0)$. This potential outcome plays a crucial role in the definition of net and mechanism effects presented below.[4] For each unit i, we observe the vector (T_i, Y_i, S_i), where $Y_i \equiv T_i Y_i(1) + (1 - T_i) Y_i(0)$ and $S_i = T_i S_i(1) + (1 - T_i) S_i(0)$. To simplify notation, in the rest of the paper we write the subscript i only when necessary. As usual in the program evaluation literature, we focus on average causal effects. The population average treatment effect is given by $ATE = E[Y(1) - Y(0)]$.[5]

Using the potential outcome $Y(1, S(0))$, the ATE can be decomposed as (e.g., Robins and Greenland, 1992; Pearl, 2001):

$$ATE = E[Y(1) - Y(1, S(0))] + E[Y(1, S(0)) - Y(0)].$$

Define the (causal) net average treatment effect or $NATE$ as:

$$NATE = E[Y(1, S(0)) - Y(0)] \tag{1}$$

and the (causal) mechanism average treatment effect or $MATE$ as:

$$MATE = E[Y(1) - Y(1, S(0))]. \tag{2}$$

The parameters $NATE$ and $MATE$ are not new in the literature, although they have received different names. $NATE$ and $MATE$ are also called the (average) pure direct and indirect effects (Robins and Greenland, 1992; Robins, 2003), or the (average) natural direct and indirect effects (Pearl, 2001). $MATE$ is also called the average causal mediation effect (Imai et al., 2010).

[4] Another potential outcome is $Y_i(0, S_i(1))$, the outcome an individual would obtain when the treatment is not given to her but she receives a value of the post-treatment variable equal to $S_i(1)$. A similar decomposition as the one to be presented below is possible using this potential outcome. If interest lies in such decomposition, the methods presented in this paper can also be applied there.

[5] We adopt the stable unit treatment value assumption (SUTVA) following Rubin (1980). This assumption is common throughout the literature, and it implies that the treatment effects at the individual level are not affected either by the method used to assign the treatment or by the treatment received by other units. In practice, this assumption rules out general equilibrium effects of the treatment that may impact individuals.

An intuitive way to think about $NATE$ is to consider $Y(1, S(0))$ as the potential outcome of an alternative counterfactual experiment in which the treatment is the same as the original one but blocks the effect of T on S by holding S fixed at $S_i(0)$ for each individual i. The net treatment effect for individual i is then the difference between the outcome of this alternative treatment, $Y_i(1, S_i(0))$, and $Y_i(0)$ from the original control treatment. An important property of $NATE$ is that it includes the part of the ATE that is totally unrelated to the mechanism variable S and also the part of the ATE that results from a change *in the way* S affects Y. That is, even though the level of S is held fixed at $S(0)$, the treatment may still affect the way in which S affects the outcome, and this is counted as part of $NATE$.[6] Also, note that $NATE$ equals zero when all the effect of T on Y works through S, and it equals the ATE when none of the effect works through S (either because T does not affect S or S does not affect Y).[7]

There are other definitions of net or direct effects available in the literature. Mealli and Rubin (2003) and Rubin (2004, 2005) define the concepts of direct and indirect effects using principal stratification (Frangakis and Rubin, 2002) as a comparison of $Y(1)$ and $Y(0)$ within the strata for which the treatment does not affect the mechanism, so that $S(0) = S(1) = s$. This parameter is typically referred to as the principal strata average direct effect or $PSDE$ (VanderWeele, 2008; Robins et al., 2007). The $PSDE$ is a special case of $NATE$ defined for the subpopulation with $S(0) = S(1) = s$, since in this strata $Y(1) = Y(1, S(0))$. It does not equal $NATE$ in (1) unless, for instance, the individual net treatment effects $Y_i(1, S_i(0)) - Y_i(0)$ are constant over the population. The parameter considered by Lee (2009) and Zhang et al. (2008) is an example of a $PSDE$, since they focus on the ATE of training on wages for those individuals who would be employed whether trained or not. Another parameter used in the literature is the average controlled direct effect or $ACDE$ (Robins and Greenland, 1992; Pearl, 2001). The $ACDE$ at a specific value $\overline{s}$ of S can be written as $ACDE = E[Y(1, S(1) = \overline{s}) - Y(0, S(0) = \overline{s})]$. The $ACDE$ gives the average difference between the counterfactual outcome under the two treatment arms controlling for the value of the mechanism variable at $\overline{s}$. For our purposes, this parameter has some undesirable features, such as not decomposing the ATE into a net and a mechanism effect[8] and that, even if in fact the treatment does not affect the mechanism variable S, the ATE can be different from the $ACDE$ if there is heterogeneity in the effect of T on Y along the values of S.

[6] This is important from a policy perspective since a policy maker typically has some degree of control over S, while very rarely over how S affects Y. Defining $NATE$ in this way is consistent with Holland's (1986) notion of a "treatment" being an intervention that can be potentially applied to each individual.

[7] For further discussion on the definitions of $NATE$ and $MATE$ see, for instance, Pearl (2001) or Flores and Flores-Lagunes (2009).

[8] For example, we could write the ATE as: $ATE = E[Y(1, S(1)) - Y(1, S(1) = \overline{s})] + ACDE + E[Y(0, S(0) = \overline{s}) - Y(0, S(0))]$. The first term gives the average effect of giving the treatment to the individuals and moving the value of the post-treatment variable from $\overline{s}$ to $S(1)$. The second term represents the average effect of giving the control treatment to the individuals and moving the value of the post-treatment variable from $S(0)$ to $\overline{s}$. These two effects are hard to interpret as "mechanism effects".

Recently there has been substantial interest on estimation of net or direct effects, mostly in fields different from economics. Much of this work has focused on point estimation of the different effects discussed above (e.g., Robins and Greenland, 1992; Pearl, 2001; Petersen et al., 2006; Gallop et al., 2009; Flores and Flores-Lagunes, 2009; Imai et al., 2010). Except for a few papers (e.g., Gallop et al., 2009), all of them require the mechanism variable to be exogenous or random after conditioning on a set of covariates, and in many cases they also require a "no-interaction" assumption.[9,10] These are strong assumptions that may not hold in typical applications in economics.

Motivated by the difficulty in point estimating these effects, others have focused on deriving bounds instead. Kaufman et al. (2005) and Cai et al. (2008) provide nonparametric bounds for the $ACDE$. The latter paper extends the former by applying the symbolic Balke-Pearl (1997) linear programming method to derive closed-form formulas for the bounds. Sjölander (2009) derives bounds for $NATE$ (or the average natural direct effect) using the same approach and assumptions as in Cai et al. (2008). As in our case, Sjölander (2009) focuses on the case in which the treatment is randomly assigned and both T and S are binary. However, he also restricts the outcome to be binary. In addition, he imposes individual-level monotonicity assumptions about the effects of (i) the treatment on the mechanism variable; (ii) the mechanism variable on the outcome; and, (iii) the treatment on the outcome. The bounds for $NATE$ derived in the following section improve those in Sjölander (2009) in several ways. First, in section 3.2 below we derive sharper bounds than those in Sjölander (2009) under similar but weaker assumptions. Second, our bounds do not require the outcome to have a bounded support. Finally, we derive the bounds analytically, as opposed to doing so by computationally solving a linear programming problem. This allows us to weaken his assumptions (section 3.2) and to consider alternative ones (section 3.3).

3 Nonparametric Partial Identification of NATE and MATE

This section presents the main results of the paper. We focus most of our discussion on $NATE$ in (1) since by definition $MATE = ATE - NATE$. We employ the principal stratification framework in our analysis (Frangakis and Rubin, 2002). The basic principal stratification with respect to a post-treatment variable S is a partition of individuals into groups such that, within each group, all individuals have the same vector $\{S(0) = s_0, S(1) = s_1\}$, where s_0 and s_1 are generic values of $S(0)$ and $S(1)$, respectively. A principal effect with respect to a principal strata is then defined as a comparison of potential outcomes within that strata.

[9] For instance, Robins and Greenland (1992) assume that for all units the effect on the outcome to a change on the treatment does not depend on the level at which the intermediate or mechanism variable is held. For a discussion of similar assumptions used in this literature, see Petersen et al. (2006).

[10] Gallop et al. (2009) avoid both types of assumptions by imposing strong parametric assumptions within a Bayesian framework, and by focusing on the $PSDE$.

Since principal strata are not affected by treatment assignment, individuals in that group are comparable and thus principal effects are causal effects.[11]

There are two main challenges for identification of $NATE$. First, the key potential outcome needed for identification of $NATE$, $Y(1, S(0))$, is generally not observed. This is in contrast to the case of estimation of the ATE, where only one of the relevant potential outcomes is missing for every unit.[12] Second, for each unit under study only one of the potential values of the mechanism variable is observed: S represents $S(1)$ for treated units and $S(0)$ for controls units. This implies that the principal strata $\{S(0) = s_0, S(1) = s_1\}$ to which each individual belongs to is not observable.[13]

Our first result is the observation that the data (T_i, Y_i, S_i) for $i = 1, \ldots, n$, which is of the kind typically available to researchers, contains information on the key potential outcome $Y(1, S(0))$ only for a particular subpopulation: those for which the treatment does not affect the mechanism. For this subpopulation we have $S_i(1) = S_i(0)$, which implies that $Y_i(1, S_i(0)) = Y_i(1)$ and, hence, $Y_i = Y_i(1, S_i(0))$ for those receiving treatment. We state this as a result in order to highlight its importance.

Result 1 *The observed data (T_i, Y_i, S_i) for $i = 1, \ldots, n$ contains information on $Y(1, S(0))$ only for those units that receive the treatment and for which the treatment does not affect the mechanism variable, so that $S_i(1) = S_i(0)$ and $Y_i = Y_i(1, S_i(0))$.*

This result does not depend on the assignment mechanism of the treatment, or on whether the mechanism variable S is binary or continuous. It implies that, under heterogeneous effects, point estimation of average net effects for other subpopulations (including the entire population) can only be based on extrapolations of $Y(1, S(0))$ to those units for which the treatment affects the mechanism, since the data contains no information on their potential outcome $Y(1, S(0))$. This result simply exemplifies the difficulty of estimating $NATE$ and $MATE$ with the data typically available: intuitively, we want to learn about a different treatment—one that holds the value of S fixed at $S(0)$—from the one at hand.

From this point on, we restrict S to be binary so that $S_i(\tau) = \{0, 1\}$ for $\tau = 0, 1$.

[11] Principal stratification generalizes the work by Imbens and Angrist (1994) and Angrist, Imbens and Rubin (1996) on the local average treatment effect interpretation of instrumental variables. For example, note that the group of "compliers" in the last two papers is the set of individuals that always comply with their treatment assignment regardless of whether their assignment is to treatment ($T = 1$) or control group ($T = 0$). Therefore, for this group we have $\{S(0) = 0, S(1) = 1\}$, where S in this case is an indicator for the actual treatment received.

[12] This implies, for instance, that even if all explanatory variables in the regression $Y = a + bT + cS + d'X + u$ were uncorrelated to the error term u (with X being a set of covariates), b does not equal $NATE$. In this simple example, the coefficient b gives the effect of T on Y holding S fixed at an arbitrary value $\overline{s}$ (i.e., the $ACDE$), and not at $S(0)$ as required by $NATE$.

[13] Note that S can be regarded as an outcome, and thus the distribution of the principal strata equals the joint distribution of the potential outcomes $\{S(1), S(0)\}$, which is not easily identifiable (e.g., Heckman, Smith and Clements, 1997).

This set up gives rise to four principal strata that are analogous to the "compliance types" of Angrist, Imbens and Rubin (1996). The four strata are given by $\{S_i(0) = 0, S_i(1) = 0\}$, $\{S_i(0) = 0, S_i(1) = 1\}$, $\{S_i(0) = 1, S_i(1) = 0\}$ and $\{S_i(0) = 1, S_i(1) = 1\}$. We refer to each of these strata as the not-affected at 0 ($n0$), the affected positively (ap), the affected negatively (an) and the not-affected at 1 ($n1$), respectively.

In what follows it is important to define the "local" $NATE$, or $LNATE$, as the net average treatment effect for a given strata:

$$LNATE_k = E[Y(1, S(0))|k] - E[Y(0)|k], \text{ for } k = n0, n1, ap, an \tag{3}$$

Although our ultimate goal is the derivation of bounds for the population $NATE$, it is important to consider the $LNATE$s since comparisons of potential outcomes within strata are causal, and the population $NATE$ is ultimately a function of the different $LNATE$s. Similarly, it is helpful to define the "local" $MATE$, or $LMATE$, as the mechanism average treatment effect for a given strata: $LMATE_k = E[Y(1)|k] - E[Y(1, S(0))|k]$, for $k = n0, n1, ap, an$. Note that $LMATE_{n0} = LMATE_{n1} = 0$.

In the next subsection we derive bounds for $LNATE_{n0}$ and $LNATE_{n1}$, which correspond to the $LNATE$s of the strata for which the treatment does not affect the mechanism. These parameters are important in their own right for several reasons. First, given Result 1, partial identification of these parameters requires less assumptions than partial identification of the population $NATE$. Second, the strata $n0$ and $n1$ can represent a large fraction of the overall population. For instance, in our empirical application, they are estimated to account for 79% of the population. Third, they can be helpful in cases where one is interested in learning whether the average net effect is different from zero (i.e., whether T has an effect on Y that is not through S) at least for a subpopulation. This is important, for instance, in the context of testing implications of the exclusion restriction assumption in just-identified instrumental variable models in the presence of heterogeneous effects (Flores and Flores-Lagunes, 2010). Finally, the bounds for $LNATE_{n0}$ and $LNATE_{n1}$ derived in the following subsection are essential for deriving bounds on the population $NATE$ in the subsequent subsections.

3.1 Basic Assumptions and Bounds on $LNATE_{n0}$ and $LNATE_{n1}$

The approach followed in this subsection is close to previous work by Zhang et al. (2008) and Lee (2009). The two assumptions presented below have also been used in the net or direct effect literature for deriving bounds on $NATE$ (Sjölander, 2009) and other direct effects (Kaufman et al., 2005; Cai et al., 2008), as well as in other settings (e.g., Imbens and Angrist, 1994; Zhang and Rubin, 2003).

First, we assume the treatment is randomly assigned, and thus the treatment received by each individual is independent of her potential outcomes and potential values of the mechanism

variable:

Assumption A1 (*Randomly Assigned Treatment).* $Y(1), Y(0), Y(1, S(0)), S(1), S(0) \perp T$.

Note that random assignment allows point identification of $E[Y(1)]$, $E[Y(0)]$, $E[S(1)]$ and $E[S(0)]$, but not $E[Y(1, S(0))]$.

Partial identification of $LNATE_{n0}$ and $LNATE_{n1}$ is complicated by the fact that the principal strata is not directly observed. Instead, we observe groups defined by the values of T_i and S_i, which contain a mix of the principal strata:

Table 1

		T_i = 0	T_i = 1
S_i	0	$ap, n0$	$an, n0$
	1	$n1, an$	$n1, ap$

A common assumption that allows identification of certain principal strata is an individual-level monotonicity assumption:

Assumption A2 (*Individual-Level Monotonicity of* T on S). $S_i(1) \geq S_i(0)$ *for all* i.

Assumption A2 states that the effect of the treatment on the mechanism variable is non-decreasing for all individuals. Imbens and Angrist (1994) and Angrist, Imbens and Rubin (1996) employed an assumption of a monotone effect of the instrument on the treatment in the context of identification of average treatment effects using instrumental variables; while Zhang et al. (2008) and Lee (2009) used a monotonicity assumption on how the treatment affected selection into employment. Here, monotonicity is applied to the effect that the treatment has on the value of the mechanism variable. In what follows, we work explicitly with the assumption of a non-negative effect of T on S. In section 3.4 we discuss the case when $S_i(1) \leq S_i(0)$ for all i is assumed instead.

Assumption A2 rules out the existence of the an principal strata, thereby allowing the identification of members of the subpopulations of $n0$ and $n1$: those units with $(T_i, S_i) = (1, 0)$ belong to the $n0$ strata, and those with $(T_i, S_i) = (0, 1)$ belong to the $n1$ strata. Therefore, we have that $E[Y(0)|n1] = E[Y|T = 0, S = 1]$ and $E[Y(1)|n0] = E[Y|T = 1, S = 0]$. Moreover, under Assumptions A1 and A2 the proportions of each of the strata in the population are point identified. Let π_{n0}, π_{n1}, π_{ap}, and π_{an} be the population proportions of each of the principal strata $n0$, $n1$, ap, and an, respectively, and also let $p_{s|t} \equiv \Pr(S_i = s|T_i = t)$ for $t, s = 0, 1$. Then, we have that $\pi_{n0} = p_{0|1}$, $\pi_{n1} = p_{1|0}$, $\pi_{ap} = p_{1|1} - p_{1|0} = p_{0|0} - p_{0|1}$ and $\pi_{an} = 0$.

From (3), note that one of the terms in each of $LNATE_{n0}$ and $LNATE_{n1}$ is point identified, while the other is not. To derive bounds for these effects, we construct bounds for these missing terms. Consider constructing bounds for $LNATE_{n0}$. In this case, $E[Y(0)|n0]$ is not point

identified because the $n0$ controls are mixed with the ap controls in the group with $T = 0$ and $S = 0$. Note that the average outcome for the individuals in this group can be written as:

$$E[Y|T = 0, S = 0] = \frac{\pi_{n0}}{\pi_{n0} + \pi_{ap}} \cdot E[Y(0)|n0] + \frac{\pi_{ap}}{\pi_{n0} + \pi_{ap}} \cdot E[Y(0)|ap] \tag{4}$$

The proportion of $n0$ in the observed group $(T, S) = (0, 0)$ can be point identified as $\pi_{n0}/(\pi_{n0} + \pi_{ap}) = p_{0|1}/p_{0|0}$. Therefore, $E[Y(0)|n0]$ can be bounded from above by the expected value of Y for the $p_{0|1}/p_{0|0}$ fraction of *largest values* of Y for those in the observed group with $T = 0$ and $S = 0$. Similarly, it can be bounded from below by the expected value of Y for the $p_{0|1}/p_{0|0}$ fraction of *smallest values* of Y for those in the same observed group.

We can follow the same approach to bound $E[Y(1)|n1]$ and derive bounds for $LNATE_{n1}$ by noting that:

$$E[Y|T = 1, S = 1] = \frac{\pi_{n1}}{\pi_{n1} + \pi_{ap}} \cdot E[Y(1)|n1] + \frac{\pi_{ap}}{\pi_{n1} + \pi_{ap}} \cdot E[Y(1)|ap] \tag{5}$$

It is also possible to construct bounds for $E[Y(0)|ap]$ and $E[Y(1)|ap]$ based on equations (4) and (5). However, it follows from Result 1 that the data contains no information on $E[Y(1, S(0))|ap]$. Therefore, $LNATE_{ap}$ is not partially identified without additional assumptions. The same holds for the population $NATE$ and $MATE$.[14]

Let y_r^{ts} be the r-th quantile of Y conditional on $T = t$ and $S = s$, or $y_r^{ts} = F^{-1}_{Y|T=t,S=s}(r)$, with $F_{\cdot}(\cdot)$ the conditional density of Y given $T = t$ and $S = s$. For example, y_r^{00} is the r-th quantile of Y conditional on $T = 0$ and $S = 0$. The bounds for $LNATE_{n0}$ and $LNATE_{n1}$, as well as for other relevant objects, are given in the following proposition.

Proposition 1 *If Assumptions A1 and A2 hold, then* $L^{n0} \leq LNATE_{n0} \leq U^{n0}$ *and* $L^{n1} \leq LNATE_{n1} \leq U^{n1}$*; where*

$$\begin{aligned}
L^{n0} &= E[Y|T = 1, S = 0] - U^{0,n0} \\
U^{n0} &= E[Y|T = 1, S = 0] - L^{0,n0} \\
L^{0,n0} &= E[Y|T = 0, S = 0, Y \leq y^{00}_{(p_{0|1}/p_{0|0})}] \\
U^{0,n0} &= E[Y|T = 0, S = 0, Y \geq y^{00}_{1-(p_{0|1}/p_{0|0})}] \\
L^{n1} &= L^{1,n1} - E[Y|T = 0, S = 1] \\
U^{n1} &= U^{1,n1} - E[Y|T = 0, S = 1] \\
L^{1,n1} &= E[Y|T = 1, S = 1, Y \leq y^{11}_{(p_{1|0}/p_{1|1})}] \\
U^{1,n1} &= E[Y|T = 1, S = 1, Y \geq y^{11}_{1-(p_{1|0}/p_{1|1})}]
\end{aligned}$$

[14] As discussed later in section 3.4, if the support of $Y(1, S(0))$ is bounded, it is possible to construct bounds for $LNATE_{ap}$, $NATE$ and $MATE$ under Assumptions A1 and A2.

Furthermore, we have: $L^{0,n0} \leq E[Y(0)|n0] \leq U^{0,n0}$, $L^{1,n1} \leq E[Y(1)|n1] \leq U^{1,n1}$, $L^{0,ap} \leq E[Y(0)|ap] \leq U^{0,ap}$, $L^{1,ap} \leq E[Y(1)|ap] \leq U^{1,ap}$; *where*

$$\begin{aligned} L^{0,ap} &= E[Y|T=0,S_i=0,Y \leq y^{00}_{1-(p_{0|1}/p_{0|0})}] \\ U^{0,ap} &= E[Y|T=0,S=0,Y \geq y^{00}_{(p_{0|1}/p_{0|0})}] \\ L^{1,ap} &= E[Y|T=1,S=1,Y \leq y^{11}_{1-(p_{1|0}/p_{1|1})}] \\ U^{1,ap} &= E[Y|T=1,S=1,Y \geq y^{11}_{(p_{1|0}/p_{1|1})}] \end{aligned}$$

Based on Proposition 1 it is possible to construct bounds for the local $NATE$ of the entire subpopulation for which $S(0) = S(1)$, which equals the $PSDE$ in Mealli and Rubin (2003) and Rubin (2004).[15] We also note that the bounds for $LNATE_{n1}$ in Proposition 1 correspond to those previously derived by Zhang et al. (2008) and Lee (2009) in a different setting.

One limitation of $LNATE_{n0}$ and $LNATE_{n1}$ for studying the part of the average effect of a treatment on an outcome that is due to a mechanism S is that they do not decompose the population ATE into a net and mechanism effect without further strong assumptions, such as requiring constant individual net treatment effects.[16] However, the bounds derived in this subsection are the basis for constructing bounds on the population $NATE$ and $MATE$ in the following two subsections.

3.2 Weak Monotonicity of Mean Potential Outcomes within Strata

We first motivate the general approach we follow to construct bounds on $NATE$. Although $E[Y(1,S(0))]$ in equation (1) is not identified from the data, note that under Assumptions A1 and A2 we can write it as a function of the expectation of $Y(1,S(0))$ in each of the strata as $E[Y(1,S(0))] = \pi_{n0}E[Y(1)|n0] + \pi_{n1}E[Y(1)|n1] + \pi_{ap}E[Y(1,S(0))|ap]$. From the previous section, all the proportions and $E[Y(1)|n0]$ are point identified, while $E[Y(1)|n1]$ is partially identified. Since the data contains no information on $Y(1,S(0))$ for the ap strata, we need to impose conditions in order to partially identify $E[Y(1,S(0))|ap]$ and construct bounds for $NATE$.

More generally, we write $NATE$ in different forms as a function of terms that are point or

[15] In our setting, $LNATE_{n0,n1} = PSDE = [\pi_{n0}/(\pi_{n0}+\pi_{n1})]LNATE_{n0} + [\pi_{n1}/(\pi_{n0}+\pi_{n1})]LNATE_{n1}$.

[16] As discussed in Flores and Flores-Lagunes (2009), this assumption is weaker than assuming a constant individual (total) effect of the treatment on the outcome. It allows for heterogeneous effects of the treatment on the outcome, but such heterogeneity is restricted to work through the mechanism S, i.e., it allows heterogeneous individual mechanism treatment effects. Nevertheless, this assumption may still be too strong in many empirical settings.

partially identified under Assumptions A1 and A2 as:

$$
\begin{aligned}
& NATE \\
= \; & E\left[Y\left(1\right)\right] + \pi_{ap} LNATE_{ap} - \pi_{n0} E\left[Y\left(0\right)|n0\right] - \pi_{n1} E\left[Y\left(0\right)|n1\right] - \pi_{ap} E\left[Y\left(1\right)|ap\right] \quad (6) \\
= \; & \pi_{n1} E\left[Y\left(1\right)|n1\right] + \pi_{n0} E\left[Y\left(1\right)|n0\right] + \pi_{ap} E\left[Y\left(1, S\left(0\right)\right)|ap\right] - E\left[Y\left(0\right)\right] \quad (7) \\
= \; & E\left[Y\left(1\right)\right] - E\left[Y\left(0\right)\right] - \pi_{ap} LMATE_{ap} \quad (8) \\
= \; & \pi_{n1} LNATE_{n1} + \pi_{n0} LNATE_{n0} + \pi_{ap} LNATE_{ap} \quad (9)
\end{aligned}
$$

It is useful to write $NATE$ in these different forms because, depending on the assumptions we impose, each of the equations above may generate different bounds, as shown below. Equations (6) and (8) add and subtract $E\left[Y\left(1\right)|ap\right]$ to $NATE$ to exploit the fact that $E\left[Y\left(1\right)\right]$ is point identified. The first two equations use the fact that either $E\left[Y\left(1\right)\right]$ or $E\left[Y\left(0\right)\right]$ is point identified and work with the remaining terms, some of which are point identified. Equation (8) is very intuitive and exploits the fact that the ATE is point identified. Remember that $NATE = ATE - MATE$. Since by definition $LMATE_{n0} = LMATE_{n1} = 0$, then $MATE = \pi_{ap} LMATE_{ap}$. The last equation writes $NATE$ as the weighted average of the $LNATEs$ of each of the strata in the population.[17] The approach we follow to derive bounds on $NATE$ consists on first obtaining bounds for the partially identified terms in equations (6) through (9). Then, we plug these bounds into those four equations, compare the resulting bounds, and keep only the lower (and upper) bounds that are not always less (greater) than another one.

We first consider assumptions analogous to those previously used in the statistics literature to bound $NATE$. In a setting like ours but with a binary outcome Sjölander (2009) assumes, in addition to Assumptions A1 and A2, that (i) $Y_i\left(1, s\right) \geq Y_i\left(0, s\right)$ for all i and all values s, and (ii) $Y_i\left(t, 1\right) \geq Y_i\left(t, 0\right)$ for all i and all values t. These assumptions imply that the individual net and mechanism treatment effects are non-negative for all the individuals in the population.[18] Assumptions similar to those in Sjölander (2009) have also been considered in the econometrics literature in other contexts. For instance, Manski (1997) and Manski and Pepper (2000) study the identifying power of the "monotone treatment response" assumption to learn about treatment responses. Their assumption states that the individual potential outcomes are a monotone function of the treatment: $Y_i\left(1\right) \geq Y_i\left(0\right)$ for all i.

Given that all the terms that are not point identified in equations (6)-(9) involve averages of

[17] Note that in equations (6)-(9) it does not make a difference if we write separately each of the terms in $LNATE_{n1}$ and $LNATE_{n0}$ or not, since for both local effects one of their terms is point identified (see equation 3). However, it is better not to brake up $LNATE_{ap}$ and $LMATE_{ap}$ into each of their terms because none of them is point identified, and we impose some of the assumptions below directly on $LNATE_{ap}$ and $LMATE_{ap}$.

[18] To see this, note that the second argument of Y represents a specific value of S and thus: $Y\left(0\right) = Y\left(0, 0\right)$ and $Y\left(1\right) = Y\left(1, S\left(0\right)\right) = Y\left(1, 0\right)$ for the $n0$ strata; $Y\left(0\right) = Y\left(0, 1\right)$ and $Y\left(1\right) = Y\left(1, S\left(0\right)\right) = Y\left(1, 1\right)$ for the $n1$ strata; and $Y\left(0\right) = Y\left(0, 0\right)$, $Y\left(1\right) = Y\left(1, 1\right)$ and $Y\left(1, S\left(0\right)\right) = Y\left(1, 0\right)$ for the ap strata.

potential outcomes for specific strata, all we need for partial identification of $NATE$ are weak mean inequalities at the principal-strata level. Thus, we employ the following assumption.

Assumption B. (*Weak Monotonicity of Mean Potential Outcomes Within Strata).*
B1. $E[Y(1)|ap] \geq E[Y(1, S(0))|ap]$. *B2.* $E[Y(1, S(0))|k] \geq E[Y(0)|k]$, for $k = n0, n1, ap$.

Assumption B provides a lower and an upper bound for $E[Y(1, S(0))|ap]$, so now equations (6)-(9) can be used to derive bounds for $NATE$. Assumption B1 implies that $LMATE_{ap} \geq 0$, and hence $MATE \geq 0$. When combined with Assumption A2, it implies the mechanism S has a non-negative average effect on Y. Similarly, Assumption B2 implies $LNATE \geq 0$ for all strata, so that $NATE \geq 0$.[19] Hence, using the fact that $ATE = NATE + MATE$, Assumption B directly implies that a lower bound for $NATE$ is 0, and an upper bound is the ATE.

Assumption B is weaker than the assumptions in Sjölander (2009) in two important ways. First, it does not require monotonicity at the individual level. This distinction is important as it may increase the plausibility of the assumption in practice by allowing the net and mechanism effects of some individuals to be negative. Second, it places conditions only on the relevant potential outcomes $Y(0)$, $Y(1)$ and $Y(1, S(0))$, and not on all possible "counterfactual" outcomes $Y(t, s)$ for all t and s.

Although assuming that $LNATE_{n0}$ and $LNATE_{n1}$ are non-negative is not strictly necessary to derive bounds on $NATE$, it is helpful in tightening the bounds. For instance, combining the result from Proposition 1 with Assumption B2, the lower bound for $LNATE_{n0}$ is now the maximum of 0 and L^{n0}. This result further implies that the upper bound for $E[Y(0)|n0]$ is the minimum of $E[Y|T = 1, S = 0]$ and $U^{0,n0}$ (see equation 3).

The following proposition presents the bounds for $NATE$ and $MATE$, as well as the local effects and relevant mean potential outcomes, under Assumptions A1, A2 and B.

Proposition 2 *If Assumptions A1, A2 and B hold, then* $\max\{0, L^{n0}\} \leq LNATE_{n0} \leq U^{n0}$, $\max\{0, L^{n1}\} \leq LNATE_{n1} \leq U^{n1}$, $0 \leq LNATE_{ap} \leq (U^{1,ap} - L^{0,ap})$, $0 \leq LMATE_{ap} \leq (U^{1,ap} - L^{0,ap})$, $\max\{L^1, L^2, L^3, L^4\} \leq NATE \leq (E[Y|T = 1] - E[Y|T = 0])$, and $0 \leq$

[19] Note that, since for the $n0$ and $n1$ strata we have that $E[Y(1)] = E[Y(1, S(0))]$, Assumption B2 implies that the local average treatment effect for these two stratas is non-negative.

$MATE \leq (E[Y|T=1] - E[Y|T=0] - \max\{L^1, L^2, L^3, L^4\})$; where

$$
\begin{aligned}
L^1 &= E[Y|T=1] - p_{0|1} \min\{E[Y|T=1,S=0], U^{0,n0}\} - p_{1|0} E[Y|T=0,S=1] \\
&\quad - (p_{1|1} - p_{1|0}) U^{1,ap} \\
L^2 &= p_{1|0} \max\{E[Y|T=0,S=1], L^{1,n1}\} + p_{0|1} E[Y|T=1,S=0] + (p_{1|1} - p_{1|0}) L^{0,ap} \\
&\quad - E[Y|T=0] \\
L^3 &= E[Y|T=1] - E[Y|T=0] - (p_{1|1} - p_{1|0})(U^{1,ap} - L^{0,ap}) \\
L^4 &= p_{1|0} \max\{0, L^{n1}\} + p_{0|1} \max\{0, L^{n0}\}
\end{aligned}
$$

Furthermore, we have: $L^{0,n0} \leq E[Y(0)|n0] \leq \min\{E[Y|T=1,S=0], U^{0,n0}\}$, $\max\{E[Y|T=0,S=1], L^{1,n1}\} \leq E[Y(1)|n1] \leq U^{1,n1}$, $L^{0,ap} \leq E[Y(0)|ap] \leq \min\{U^{0,ap}, U^{1,ap}\}$, $\max\{L^{0,ap}, L^{1,ap}\} \leq E[Y(1)|ap] \leq U^{1,ap}$ *and* $L^{0,ap} \leq E[Y(1, S(0))|ap] \leq U^{1,ap}$.

Proposition 2 states that under Assumptions A1, A2 and B the upper bound for $NATE$ equals the estimated ATE, so that the lower bound for $MATE$ is zero. This particular upper bound for $NATE$ comes from equation (8), and it is always less or equal than the other three upper bounds derived using equations (6), (7) and (9). However, the lower bound for $NATE$ is the maximum of the bounds derived from each of the four equations (6) to (9). Depending on the data, Proposition 2 implies that it is possible to obtain a lower (upper) bound for $NATE$ ($MATE$) that is above zero (below the estimated ATE).

The bounds in Proposition 2 extend those in Sjölander (2009) by allowing for an outcome with unbounded support. Moreover, we tighten the bounds derived in Sjölander (2009) by using the trimming procedure to derive bounds in section 3.1.[20]

3.3 Weak Monotonicity of Mean Potential Outcomes across Strata

A potentially unattractive feature of Assumption B is that it imposes restrictions on the sign of the effects of interest. In this subsection we consider assumptions stating that mean potential outcomes vary weakly monotonically across strata. To the best of our knowledge, the assumptions presented in this subsection have not been considered before to derive bounds on net and mechanism effects, although similar assumptions have been previously used in other settings. For instance, Manski and Pepper (2000) introduce a "monotone instrumental variable" assumption for identification of treatment effects, which states that mean responses vary weakly monotonically across subpopulations defined by specific values of the instrument.[21]

[20] In particular, if we restrict the outcome to be binary and set $L^{0,n0} = L^{1,n1} = L^{0,ap} = L^{1,ap} = 0$, $U^{0,n0} = U^{1,n1} = U^{0,ap} = U^{1,ap} = 1$, $\max\{0, L^{n1}\} = 0$ and $\max\{0, L^{n0}\} = 0$, we obtain the bounds in Sjölander (2009) (see equation (14) in that paper).

[21] For example, in the problem of estimating the effect of attending college on future earnings using measured ability as a "monotone instrument", this assumption states that individuals with higher measured ability have weakly higher mean potential future earnings than those with lower measured ability.

Our assumptions, however, condition on the basic principal strata, i.e., on specific values of $S(0)$ and $S(1)$. Another example of this type of assumptions appears in Zhang et al. (2008), who assume that the mean potential wages of those individuals who would be employed whether they attended training or not are always greater than or equal to those of the individuals who would be employed when trained but unemployed when not trained. They use this assumption to tighten the bounds for the average treatment effect (ATE) of training on wages for the subpopulation of individuals who would be employed whether they received training or not.[22] Formally, our assumption is:

Assumption C. (*Weak Monotonicity of Mean Potential Outcomes Across Strata).*
C1. $E[Y(1,S(0))|ap] \geq E[Y(1)|n0]$. *C2.* $E[Y(1)|n1] \geq E[Y(1,S(0))|ap]$. *C3.* $E[Y(0)|ap] \geq E[Y(0)|n0]$. *C4.* $E[Y(0)|n1] \geq E[Y(0)|ap]$. *C5.* $E[Y(1)|ap] \geq E[Y(1)|n0]$. *C6.* $E[Y(1)|n1] \geq E[Y(1)|ap]$.

Assumption C states that the mean potential outcomes of those who receive a value of the mechanism equal to one if treated and zero if not are less (greater) than or equal to the corresponding mean potential outcomes of those who always receive a value of the mechanism variable of one (zero). For example, consider the empirical application presented in the following section, where we study what part of the effect of a training program is due to the obtainment of a high school, GED, or vocational degree. In this case, Assumption C states that the mean potential outcomes of those who receive a degree only if trained are less (greater) than or equal to the corresponding mean potential outcomes of those who always (never) receive a degree whether trained or not. Assumption C formalizes the notion that some strata have more favorable characteristics and thus better potential outcomes on average. As further discussed in the following section, the direction of the inequalities in Assumptions C1-C6 can be changed depending on the particular application.

Assumption C is likely to hold in many applications since often times we expect the potential outcomes to differ weakly monotonically across strata, or we may even have a theory that predicts so. Importantly, combining Assumptions A1, A2 and C yields some testable implications that can be used to falsify the assumptions. In particular, they imply that (see equations 4 and 5, respectively):

$$E[Y|T=0,S=1] \geq E[Y|T=0,S=0] \text{ and } E[Y|T=1,S=1] \geq E[Y|T=1,S=0] \quad (10)$$

We illustrate the use of these testable implications in the empirical application of section 4.

Assumptions C1 and C2 provide a lower and an upper bound for $E[Y(1,S(0))|ap]$, respectively. Although Assumptions C3-C6 are not strictly necessary to derive bounds for $NATE$,

[22] More generally, the assumptions below are closely related to the stochastic dominance conditions commonly used in the partial identification literature (e.g., Manski 2003, 2007).

they are helpful in tightening the bounds. To illustrate how we derive bounds under Assumptions A1, A2 and C, consider deriving bounds for $E[Y(0)|ap]$. Assumption C4 implies that an upper bound for $E[Y(0)|ap]$ is $E[Y(0)|n1] = E[Y|T=0,S=1]$, which combined with the result from Proposition 1 yields that an upper bound for $E[Y(0)|ap]$ is the minimum of $E[Y|T=0,S=1]$ and $U^{0,ap}$. Assumption C3 implies that $E[Y(0)|n0]$ is a lower bound for $E[Y(0)|ap]$, which combined with equation (4) yields $E[Y(0)|ap] \geq E[Y|T=0,S=0]$.[23] Since by definition $E[Y|T=0,S=0] \geq L^{0,ap}$, we have that the lower bound for $E[Y(0)|ap]$ is $E[Y|T=0,S=0]$. Note that following a similar argument, Assumption C3 implies that $E[Y(0)|n0] \leq E[Y|T=0,S=0]$, and combining Assumption C6 with equation (5) yields $E[Y(1)|ap] \leq E[Y|T=1,S=1]$ and $E[Y(1)|n1] \geq E[Y|T=1,S=1]$. The following proposition presents the complete set of bounds when Assumption C is added to Assumptions A1 and A2.

Proposition 3 *If Assumptions A1, A2 and C hold, then* $\overline{L}^{n0} \leq LNATE_{n0} \leq U^{n0}$, $\overline{L}^{n1} \leq LNATE_{n1} \leq U^{n1}$, $\overline{L}^{ap} \leq LNATE_{ap} \leq \overline{U}^{ap}$, $\overline{L}^{ap}_m \leq LMATE_{ap} \leq \overline{U}^{ap}_m$, $\overline{L} \leq NATE \leq \min\{\overline{U}^1, \overline{U}^2\}$, *and* $\overline{L}_m \leq MATE \leq \overline{U}_m$; *where*

$$
\begin{aligned}
\overline{L}^{n0} &= E[Y|T=1,S=0] - E[Y|T=0,S=0] \\
\overline{L}^{n1} &= E[Y|T=1,S=1] - E[Y|T=0,S=1] \\
\overline{L}^{ap} &= E[Y|T=1,S=0] - \min\left\{U^{0,ap}, E[Y|T=0,S=1]\right\} \\
\overline{U}^{ap} &= U^{1,n1} - E[Y|T=0,S=0] \\
\overline{L}^{ap}_m &= \max\left\{L^{1,ap}, E[Y|T=1,S=0]\right\} - U^{1,n1} \\
\overline{U}^{ap}_m &= E[Y|T=1,S=1] - E[Y|T=1,S=0] \\
\overline{L} &= E[Y|T=1] - E[Y|T=0] - \left(p_{1|1} - p_{1|0}\right)\overline{U}^{ap}_m \\
\overline{U}^1 &= E[Y|T=1] - E[Y|T=0] + p_{1|1}\left(U^{1,n1} - E[Y|T=1,S=1]\right) \\
\overline{U}^2 &= E[Y|T=1] - E[Y|T=0] - \left(p_{1|1} - p_{1|0}\right)\overline{L}^{ap}_m \\
\overline{L}_m &= E[Y|T=1] - E[Y|T=0] - \min\{\overline{U}^1, \overline{U}^2\} \\
\overline{U}_m &= E[Y|T=1] - E[Y|T=0] - \overline{L}
\end{aligned}
$$

Furthermore, we have: $L^{0,n0} \leq E[Y(0)|n0] \leq E[Y|T=0,S=0]$, $E[Y|T=1,S=1] \leq E[Y(1)|n1] \leq U^{1,n1}$, $E[Y|T=0,S=0] \leq E[Y(0)|ap] \leq \min\{U^{0,ap}, E[Y|T=0,S=1]\}$, $\max\{L^{1,ap}, E[Y|T=1,S=0]\} \leq E[Y(1)|ap] \leq E[Y|T=1,S=1]$ *and* $E[Y|T=1,S=0] \leq E[Y(1,S(0))|ap] \leq U^{1,n1}$.

Under the assumptions in Proposition 3, the lower bounds derived using equations (7) and (8) are equal to $\overline{L}$, and they are always greater than or equal to those derived using equations

[23] Following Zhang and Rubin (2003), note that Assumption C3 combined with equation (4) yields: $E[Y(0)|ap] = (\pi_{n0}/\pi_{n0} + \pi_{ap}) \cdot E[Y_i(0)|ap] + (\pi_{ap}/\pi_{n0} + \pi_{ap}) \cdot E[Y_i(0)|ap] \geq E[Y|T=0,S=0]$.

(6) and (9). The upper bounds $\overline{U}^1$ and $\overline{U}^2$ come from equations (7) and (8), respectively, and they are always less than or equal to those derived using equations (6) and (9).

The upper bounds for $NATE$ in Proposition 3 are always greater than or equal to the estimated ATE, and hence the lower bound for $MATE$ is always less than or equal to zero. To see this, note that by definition $U^{1,n1} \geq E[Y|T=1,S=1]$, so $\overline{U}^1 \geq E[Y|T=1] - E[Y|T=0]$. Additionally, $U^{1,n1} \geq E[Y|T=1,S=1] \geq L^{1,ap}$ and $U^{1,n1} \geq E[Y|T=1,S=1] \geq E[Y|T=1,S=0]$ (where the second inequality comes from equation 10) imply that $\overline{L}_m^{ap} \leq 0$. This, combined with the fact that $p_{1|1} - p_{1|0} = \pi_{ap} \geq 0$, implies that $\overline{U}^2 \geq E[Y|T=1] - E[Y|T=0]$.

Finally, we combine Assumptions A1, A2, B, and C. The combination of all assumptions adds the following testable implication to those presented in (10) under Assumptions A1, A2, and C:[24]

$$E[Y|T=1,S=1] \geq E[Y|T=0,S=0] \tag{11}$$

The following proposition presents the complete set of bounds under Assumptions A1, A2, B, and C.

Proposition 4 *If Assumptions A1, A2, B and C hold, then* $\max\{0, \overline{L}^{n0}\} \leq LNATE_{n0} \leq U^{n0}$, $\max\{0, \overline{L}^{n1}\} \leq LNATE_{n1} \leq U^{n1}$, $\max\{0, \overline{L}^{ap}\} \leq LNATE_{ap} \leq \widetilde{U}^{ap}$, $0 \leq LMATE_{ap} \leq \widetilde{U}_m^{ap}$, $\max\{\widetilde{L}^1, \widetilde{L}^2\} \leq NATE \leq \widetilde{U}$, and $0 \leq MATE \leq \widetilde{U}_m$; where

$$\begin{aligned}
\widetilde{U}^{ap} &= E[Y|T=1,S=1] - E[Y|T=0,S=0] \\
\widetilde{U}_m^{ap} &= E[Y|T=1,S=1] - \max\{E[Y|T=1,S=0], E[Y|T=0,S=0]\} \\
\widetilde{L}^1 &= p_{1|0} \max\{E[Y|T=1,S=1], E[Y|T=0,S=1]\} \\
&\quad + (p_{1|1} - p_{1|0}) \max\{E[Y|T=1,S=0], E[Y|T=0,S=0]\} \\
&\quad + p_{0|1} E[Y|T=1,S=0] - E[Y|T=0] \\
\widetilde{L}^2 &= p_{1|0} \max\{0, \overline{L}^{n1}\} + p_{0|1} \max\{0, \overline{L}^{n0}\} + (p_{1|1} - p_{1|0}) \max\{0, \overline{L}^{ap}\} \\
\widetilde{U} &= E[Y|T=1] - E[Y|T=0] \\
\widetilde{U}_m &= E[Y|T=1] - E[Y|T=0] - \max\{\widetilde{L}^1, \widetilde{L}^2\}
\end{aligned}$$

Furthermore: $L^{0,n0} \leq E[Y(0)|n0] \leq \min\{E[Y|T=0,S=0], E[Y|T=1,S=0]\}$, $\max\{E[Y|T=1,S=1], E[Y|T=0,S=1]\} \leq E[Y(1)|n1] \leq U^{1,n1}$, $E[Y|T=0,S=0] \leq E[Y(0)|ap] \leq \min\{U^{0,ap}, E[Y|T=0,S=1], E[Y|T=1,S=1]\}$, $\max\{L^{1,ap}, E[Y|T=1,S=0], E[Y|T=0,S=0]\} \leq E[Y(1)|ap] \leq E[Y|T=1,S=1]$ *and* $\max\{E[Y|T=1,S=0], E[Y|T=0,S=0]\} \leq E[Y(1,S(0))|ap] \leq E[Y|T=1,S=1]$.

[24] Note that Assumptions B and C imply $E[Y(1)|n1] \geq E[Y(0)|n1] \geq E[Y(0)|ap] \geq E[Y(0)|n0]$ and $E[Y(1)|ap] \geq E[Y(0)|ap] \geq E[Y(0)|n0]$. Combining these inequalities with equations (4) and (5) yields (11).

The first lower bound for $NATE$ in proposition 2 ($\widetilde{L}^1$) comes from equation (7), while the second ($\widetilde{L}^2$) comes from equation (9). As expected from Assumption B, $\widetilde{L}^2$ implies that the lower bound on $NATE$ is always greater or equal to zero, so that the upper bound on $MATE$ is always less or equal to the estimated ATE. Similar to Proposition 2, under Assumptions A1, A2, B, and C the upper bound for $NATE$ equals the estimated ATE, so the lower bound for $MATE$ equals zero. Hence, the way in which Assumption B helps tighten the bounds for our parameters when added to Assumptions A1, A2 and C is similar to the way it helps tighten the bounds when added to Assumptions A1 and A2.

3.4 Remarks

Remark 1. Propositions 1 through 4 suggest that the bounds for $NATE$ and $MATE$ will be more informative in cases where the proportion of the affected-positively strata in the population ($\pi_{ap} = p_{1|1} - p_{1\backslash 0}$) is smaller. This is intuitive since, as discussed in section 3.1, the data is not informative about the potential outcome $Y(1, S(0))$ for this strata; thus, the larger the proportion of the ap strata in the population, the less information about that potential outcome is contained in the data.

Remark 2. In some applications it may not be necessary to impose all the conditions presented in sections 3.2 and 3.3. As previously discussed, Assumptions C3-C6 and B2 for the $n0$ and $n1$ strata are not strictly necessary to derive bounds on $NATE$, so they may not be invoked in a particular application if the rest of the assumptions are enough to inform about the question of interest. To derive bounds on $NATE$, all we need are assumptions that yield bounds on $E[Y(1, S(0))|ap]$. For example, we could maintain Assumptions B1 and C1, and drop Assumptions B2 and C2. Moreover, if interest lies in obtaining only one of the bounds on $NATE$, we may drop Assumptions B1 and C2 (for a lower bound) or Assumptions B2 and C1 (for an upper bound).

Remark 3. It is possible to construct bounds for $NATE$ and $MATE$ employing only Assumptions A1 and A2 when the support of $Y(\cdot)$ is bounded. Letting $\left[Y^L, Y^U\right]$ denote the range of $Y(\cdot)$, in this case we have that $Y^L \leq E[Y(1, S(0))|ap] \leq Y^U$, and the same approach as in the previous sections can be used to derive bounds on $NATE$ and $MATE$. Note that the bounds for the rest of the means of $Y(0)$ and $Y(1)$ for all the strata in Proposition 1 fall inside the interval $\left[Y^L, Y^U\right]$, so all the bounds in Proposition 1 are unaffected.[25] Clearly, the bounds derived under Assumptions A1, A2 and the boundedness of the support of $Y(\cdot)$ are wider than those in Propositions 2 through 4.[26]

[25] The bounds for $LNATE_{ap}$, $LMATE_{ap}$, $NATE$, and $MATE$ corresponding to this case are available from the authors upon request.

[26] As shown in Sjölander (2009) for the case of a binary outcome, bounds for $NATE$ and $MATE$ can also be constructed using only Assumption A1 (Random Assignment). These bounds, however, are likely to be too wide in practice.

Remark 4. In this paper we concentrate on the case in which the effect of the treatment on the mechanism variable is assumed to be non-decreasing (Assumption A2). Although the bounds in Propositions 1 through 4 would not be the same if that effect were assumed to be non-increasing, it is straightforward to apply the same approach used in the previous sections to derive bounds under that version of Assumption A2. Assuming $S_i(1) \leq S_i(0)$ for all i (call it Assumption A2') rules out the existence of the ap strata, so $E[Y|T=0,S=0] = E[Y(0)|n0]$, $E[Y|T=1,S=1] = E[Y(1)|n1]$, and $(T_i,S_i)=(0,1)$ and $(T_i,S_i)=(1,0)$ are now a mixture of the an strata with the $n0$ and $n1$ stratas, respectively. In this case, Assumptions B and C would be about the an strata instead of the ap strata, and the direction of the inequalities may also be changed. For instance, if the mechanism is assumed to have a non-negative effect on the outcome and the $LNATE$s for all stratas are assumed to be non-negative, Assumption B could be stated as (call it Assumption B'): B1'. $E[Y(1,S(0))|an] \geq E[Y(1)|an]$. B2'. $E[Y(1,S(0))|k] \geq E[Y(0)|k]$, for $k = n0, n1, an$.[27]

Remark 5. The last three remarks suggest that the assumptions in the previous sections can be changed, and the bounds adjusted, depending on their plausibility, identifying power, and the economic theory behind any particular application. First, some particular assumptions can be dropped if they are not tenable or needed in a particular application (Remark 2). Second, the direction of the inequalities can also be changed (Remark 4). Finally, the mean potential outcomes of the strata that we use in Assumption C can be changed. For instance, if Assumption C1 is not justifiable in a particular application, it could be changed for a more conservative version requiring that $E[Y(1,S(0))|ap] \geq E[Y(0)|n0]$. This last assumption requires the mean outcome of $Y(1,S(0))$ for the ap strata to be no less than the average outcome under control for the $n0$ strata (as opposed to the average outcome under treatment for the $n0$ strata, as in Assumption C1). In sum, the specific bounds we derived in this section can be adjusted to different empirical applications.

4 Empirical Application

In this section we illustrate the identifying power of the bounds in Propositions 1 through 4 by analyzing what part of the effect of a training program on employment and weekly earnings is due to the individual's obtainment of a high school, GED, or vocational degree. The particular program we consider is Job Corps (JC), one of the largest federally-funded job training programs in the United States. It provides economically disadvantaged young people (ages 16 to 24) with academic, vocational and social skills training at over 120 centers throughout the country, where

[27] Interestingly, note that in this particular example both assumptions B1' and B2' provide a lower bound for $E[Y(1,S(0))|an]$, so under Assumptions A1, A2' and B' it is not possible to obtain an upper bound for $NATE$ without additional assumptions (such as those in Assumption C). A complete set of results under these alternative assumptions, along with the bounds resulting from adding an assumption analogous to Assumption C, is available from the authors upon request.

most participants live while enrolled. In addition, JC provides health services, a stipend during program enrollment, counseling, and job search assistance when exiting the program.

In the mid-1990s the U.S. Department of Labor funded the National Job Corps Study (NJCS), a randomized experiment to evaluate the effectiveness of JC. A random sample of all pre-screened eligible applicants in the 48 contiguous states and the District of Columbia was randomly assigned into treatment and control groups (9,409 and 5,977 individuals, respectively), with the second group being denied access to JC for three years. Both groups were tracked with a baseline interview immediately after randomization and then at 12, 30 and 48 months thereafter. The NJCS found a statistically significant positive effect of JC 12 and 16 quarters after randomization on weekly earnings ($24.5 and $25.2, respectively) and on the probability of being employed (4.4 and 3.3 percent, respectively).[28]

In this empirical application we go a step further and analyze possible mechanisms or channels through which JC affects labor outcomes by employing the bounds developed in this paper. In particular, we study what part of that positive effect is due to the completion of a high school, GED, or vocational degree, relative to other components of the program such as job search assistance, social skills training, health services, counseling, and residential living. Learning about the relative effectiveness of different types of components of the program will increase our understanding of JC and is relevant for policy purposes.

Our data comes from the NJCS, and our specific sample consists of all individuals with non-missing values on treatment status, the mechanism variable, and the outcomes considered. We focus on the outcomes measured twelve quarters after random assignment, which corresponds to the time the embargo from the program ended for the control group. The treatment and control groups employed consist of 5,045 and 2,975 individuals, respectively.[29] Since in the data there is non-compliance with the treatment assignment, the (total) average treatment effects we estimate below should be interpreted as average "intent-to-treat" effects. For consistency with the previous sections and to avoid introducing new notation, however, we refer to this effect as the ATE of the program on the outcome and to the treatment as participation in JC.[30]

Table 2 presents point estimates for some relevant parameters. The ATE of the program on the probability of being employed 12 quarters after random assignment is 4 percent, while the ATE on weekly earnings is $18. The ATE of JC on the probability of obtaining a high school, GED, or vocational degree is 21 percent. All three effects are highly statistically significant. Given the large effect of JC on the probability of obtaining a degree, one would expect this to

[28] The effects reported in the NJCS are interpreted as average effects for those individuals that comply with their treatment assignment. For further description of the JC program and the NJCS see Schochet, Burghardt and Glazerman (2001) and Flores-Lagunes, Gonzalez and Neumann (2010).

[29] In this application we abstract from the problems of sample attrition over time and missing values. Lee (2009), who employs the same data, suggests that the attrition/non-response problem is not serious.

[30] The proportion of those in the treatment group who enroll in JC was 73%, and the proportion of those in the control group that managed to enroll in JC was 1.4%.

be an important mechanism through which the program affects future labor outcomes.

The monotonicity assumption A2 states that participating in JC has a non-negative individual level effect on the obtainment of a degree, so that there are no individuals who would obtain a degree if they did not participate in JC and would not if they participate. This assumption is plausible in this setting given that JC facilitates the obtainment of such a degree. In this application, the $n0$ ($n1$) strata consists of those individuals who would never (always) obtain a degree regardless of whether they participate in JC or not; and the ap strata consists of those who would obtain a degree if they enroll in JC, but would not if they did not enroll. From Table 2, the estimated proportions of the stratas $n0$, $n1$, and ap in the population are 0.34, 0.45 and 0.21, respectively, so 79 percent of the population belong to the strata for which the treatment does not affect the mechanism variable.

Assumption A2 and B1 together imply that the obtainment of a degree has a non-negative average effect on employment and earnings, which is consistent with conventional human capital theories in economics. Assumption B2 states that the $LNATE$ for all strata are non-negative, or that the other channels have a non-negative effect on labor outcomes. Since other components of the JC program are aimed at improving the future labor outcomes of their participants (e.g., job search assistant, social skills training), we believe this assumption is likely to be satisfied.

Assumption C states that the average potential outcomes of the individuals who obtain a degree only if they participate in JC is no less (no greater) than the corresponding average potential outcomes of those who never (always) obtain a degree regardless of their participation in JC. We believe this assumption is likely to hold in our application. Although Assumption C is not directly testable, indirect evidence regarding its plausibility can be gained from comparing the baseline characteristics of the individuals in different stratas. In particular, one can check if the average baseline characteristics of the $n1$ strata are "better"—in the sense that they are related to better labor market outcomes—than those of the ap strata, and if the ap strata in turn has better average baseline characteristics than those of the $n0$ strata. For this purpose, pre-treatment values of the outcome are relevant as they are likely highly correlated with the potential outcomes in Assumption C. In our application, the probability of being employed and the average weekly earnings in the year prior to randomization of both the $n1$ and ap stratas are statistically greater than those of the $n0$ strata; while the differences of those two variables between the ap and $n1$ stratas are not statistically different from zero.[31] We interpret these results as a failure of the data to provide indirect evidence against Assumption C. The last three rows of Table 2 verify that the testable implications in (10) under Assumption A1, A2

[31] The probability of being employed in the year prior to randomization for the $n0$, ap, and $n1$ strata are, respectively (standard errors in parenthesis): 0.153 (0.36); 0.205 (0.40); 0.216 (0.41). The corresponding numbers for the average weekly earnings in the year prior to randomization are: 86.26 (107.17); 117.73 (530.33); 109.74 (112.51). The means for the $n0$ and $n1$ strata are calculated from the groups with $(T_i, S_i) = (1, 0)$ and $(T_i, S_i) = (0, 1)$, respectively. The mean for the ap strata is estimated by writing it as a function of the population mean, the means for the $n0$ and $n1$ stratas, and the strata proportions in the population.

and C, and (11) after adding Assumption B, hold in this application. Hence, our assumptions are not falsified by the data.

Table 3 and Table 4 show the estimated bounds for the employment and earnings outcomes, respectively, for each of the four propositions in Section 3. We provide standard errors for each of the bounds to give a sense of the accuracy with which they are estimated.[32] In general, the bounds in Tables 3 and 4 are precisely estimated. Since the main purpose of the application is to illustrate the identifying power of the bounds derived in the previous section, we focus our discussion below on the point estimates of the bounds and abstract from performing statistical inference.[33]

Under Assumptions A1 and A2 only bounds for $LANE_{n0}$ and $LANE_{n1}$ can be obtained. In this application, they have little identifying power for both outcomes. Adding Assumption B narrows the bounds for $LANE_{n0}$ and $LANE_{n1}$ by setting the lower bound to zero. The lower bound for both the population $NATE$ and $MATE$ is zero, while the upper bound is the estimated ATE. As discussed in Section 3.2, these particular bounds (zero and the estimated ATE) come directly from Assumption B, so the data in this application does not provide any additional information to tighten these bounds further for either of the outcomes in Tables 3 and 4. This is not the case with Assumption C.

Assumptions A1, A2 and C have more identifying power in this application. For instance, note that the lower bound on $LANE_{n1}$ suggests a positive net average treatment effect of 4.4 percent (Table 3) and \$15.4 (Table 4) for those individuals who would always obtain a degree whether trained or not. Since the obtainment of the degree is not affected by enrollment into JC for this subpopulation, this implies that there are other benefits to participating in JC besides the obtainment of a degree (at least for this subpopulation). Moreover, the lower bound of 4.4 percent for employment (\$15.4 for earnings) is obtained by setting $E[Y(1)|n1] = E[Y(1)|ap]$, so to the extent that $E[Y(1)|n1]$ is larger than $E[Y(1)|ap]$ the true $LNATE_{n1}$ will be larger than 4.4 percent (\$15.4). Also, note that this may suggest the presence of heterogeneity on the individual total treatment effect of JC on employment, since the lower bound for $LNATE_{n1}$ is slightly above the estimated (population) ATE of 4.1 percent.[34] This illustrates how knowledge

[32] The standard errors for the estimators of the bounds not involving minimum or maximum operators are obtained with 5,000 bootstrap replications. For the estimators of bounds involving those two operators, we combine the bootstrap results for the potential bounds not involving those two operators with the results from Clark (1961), who provides an algorithm to approximate the variance of the maximum of two or more random variables having a joint normal distribution. Finally, for those bounds truncated at zero we follow Cai et al. (2008) and calculate the standard errors for the estimators employing the formula for a truncated (at zero) normal distribution.

[33] We warn the reader that it is not straightforward to construct valid confidence intervals based on the standard errors reported in Tables 3 and 4. A complete analysis of inference based on the bounds presented in Propositions 1 through 4 is beyond the scope of this paper, which main focus is on identification. The interested reader is referred to recent work on inference for partially identified models defined by moment inequalities by Chernozhukov et al. (2007), Bugni (2010), Romano and Shaikh (2010) and Andrews and Soares (2010).

[34] Remember that for this subpopulation the $LNATE_{n1}$ equals the average treatment effect for this subpopu-

about the local $NATE$ for a specific strata can be helpful in practice.

Once Assumption C is added to A1 and A2, the bounds on the population $NATE$ and $MATE$ provide valuable information. In the case of employment, the lower bound on $NATE$ suggests a positive average effect of at least 1 percent of JC on employment net of its effect through the obtainment of a degree. This implies that the average effect of JC on employment that is due to the obtainment of a degree ($MATE$) is at most 3 percent, or 75 percent of the total ATE. Similarly, in the case of earnings the upper bound on $MATE$ is \$14.7, or approximately 81 percent of the total ATE. Hence, JC seems to provide other benefits to their participants besides the obtainment of a degree. These results highlight the identifying power of Assumption C.

The last vertical panel of Tables 3 and 4 show the estimated bounds under Assumptions A1, A2, B and C. As expected, adding Assumption B helps increase the lower bound, as now each of the $LNATE$s is at least zero. Similarly, now the upper bound on $NATE$ is the estimated ATE, so the lower bound on $MATE$ is zero. In the case of employment, the lower bound on $NATE$ goes from 1 percent under Assumptions A1, A2 and C, to 2 percent once Assumption B is added. This decreases the upper bound on $MATE$ from 3 to 2 percent. For earnings, the lower bound on $NATE$ is now \$6.9 while the upper bound on $MATE$ is \$11.3. Hence, under these assumptions, the average effect of JC on employment (earnings) that is due to the obtainment of a degree is at most half (sixty percent) of the total ATE of JC on the probability of employment (earnings).

We close this section by underscoring the results in Table 4, where the outcome is average weekly earnings. Despite the support of this outcome not being bounded, we obtain results that are informationally similar to those obtained with the binary employment outcome. This illustrates the usefulness of our bounds beyond settings with binary outcomes.

5 Conclusion

This paper analyzed nonparametric partial identification of net and mechanism average treatment effects ($NATE$ and $MATE$) in a heterogeneous effects setting and allowing for an outcome with unbounded support. We derive bounds for the population $NATE$ and $MATE$ within the principal stratification framework of Frangakis and Rubin (2002) by writing the $NATE$ as a function of average potential outcomes in each of the strata defined by the potential values of the mechanism variable.

Our bounds are based on two assumptions that have been previously used in the literature (Assumptions A1 and A2): random treatment assignment and individual-level monotonicity of the effect of the treatment on the mechanism. These two assumptions are combined with

lation, since $Y(1, S(0)) = Y(1)$.

one or both of two additional sets of assumptions. The first set (Assumption B) imposes weak monotonicity of potential outcomes within strata. These assumptions are weaker than similar assumptions used previously in the statistics literature, and the bounds we derive based on them are tighter and more general by not requiring an outcome with a bounded support. The second set of assumptions (Assumption C) involves weak monotonicity of potential outcomes across strata. These assumptions had not been considered before to derive bounds on $NATE$ and $MATE$, and they can have substantial identifying power, as illustrated in our empirical application. Importantly, Assumption C provides testable implications and it seems likely to hold in many economic applications.

Several extensions of the results contained here are ongoing. Extensions to cases where the treatment or the mechanism variable are multivalued are important. In such cases, the number of strata as well as the number of unidentified objects increase, reflecting the difficulty of answering the question of interest with the available data. It is also important to construct bounds for $NATE$ and $MATE$ in settings in which the treatment is not randomly assigned. Finally, derivation of bounds when an instrumental variable for the mechanism variable is available is also at the top of our research agenda.

6 Appendix

From Section 3, the relevant point identified objects in our setting are: $\pi_{n0} = p_{0|1}$, $\pi_{n1} = p_{1|0}$, $\pi_{ap} = p_{1|1} - p_{1|0} = p_{0|0} - p_{0|1}$, $E[Y(1)] = E[Y|T=1]$, $E[Y(0)] = E[Y|T=0]$, $E[Y(1)|n0] = E[Y|T=1, S=0]$, $E[Y(0)|n1] = E[Y|T=0, S=1]$, $\pi_{n0}E[Y_i(0)|n0] + \pi_{ap}E[Y_i(0)|ap] = p_{0|0}E[Y_i|T_i=0, S_i=0]$ and $\pi_{n1}E[Y_i(1)|n1] + \pi_{ap}E[Y_i(1)|ap] = p_{1|1}E[Y_i|T_i=1, S_i=1]$.

Proof of Proposition 1. It follows directly from the arguments in the text.

Proof of Proposition 2. We start by deriving bounds for the non-point identified mean potential outcomes of the stratas, and for all the local net and mechanism average treatment effects. *Bounds for* $E[Y(0)|n0]$: Ass. B2 implies $E[Y(1)|n0] = E[Y|T=1, S=0] \geq E[Y(0)|n0]$. Ass. B does not provide any additional information for a lower bound of $E[Y(0)|n0]$. Combining this with the result in Prop. 1, and since $U^{0,n0}$ can be above or below $E[Y|T=1, S=0]$, we have: $L^{0,n0} \leq E[Y(0)|n0] \leq \min\{U^{0,n0}, E[Y|T=1, S=0]\}$.[35]

Bounds for $E[Y(1)|n1]$: Ass. B2 implies $E[Y(1)|n1] \geq E[Y(0)|n1] = E[Y|T=0, S=1]$. Ass. B does not provide any additional information for an upper bound of $E[Y(1)|n1]$. Combining this with the result in Prop. 1 we have: $\max\{L^{1,n1}, E[Y|T=0, S=1]\} \leq E[Y(1)|n1] \leq U^{1,n1}$.

Bounds for $E[Y(0)|ap]$: Ass. B1 and B2 imply $E[Y(1)|ap] \geq E[Y(0)|ap]$, which combined with the results from Prop. 1 gives that $U^{1,ap}$ is another upper bound for $E[Y(0)|ap]$. Ass. B does not provide any additional information for a lower bound of $E[Y(0)|ap]$. Hence, $L^{0,ap} \leq E[Y(0)|ap] \leq \min\{U^{0,ap}, U^{1,ap}\}$.

Bounds for $E[Y(1)|ap]$: Ass. B implies $E[Y(1)|ap] \geq E[Y(0)|ap]$, which combined with the results from Prop. 1 gives that $L^{0,ap}$ is another lower bound for $E[Y(1)|ap]$. Hence, $\max\{L^{0,ap}, L^{1,ap}\} \leq E[Y(1)|ap] \leq U^{1,ap}$.

Bounds for $E[Y(1, S(0))|ap]$: Ass. B1 and B2 imply $E[Y(1)|ap] \geq E[Y(1, S(0))|ap] \geq E[Y(0)|ap]$, which combined with the results above gives $L^{0,ap} \leq E[Y(1, S(0))|ap] \leq U^{1,ap}$.

Bounds for $LNATE_{n0}$: From (3), $LNATE_{n0} = E[Y|T=1, S=0] - E[Y(0)|n0]$. Using the bounds previously derived for $E[Y(0)|n0]$ we have:[36] $\max\{0, L^{n0}\} \leq LNATE_{n0} \leq U^{n0}$.

Bounds for $LNATE_{n1}$: From (3), $LNATE_{n1} = E[Y(1)|n1] - E[Y|T=0, S=1]$. Using the bounds previously derived for $E[Y(1)|n1]$ we have: $\max\{0, L^{n1}\} \leq LNATE_{n1} \leq U^{n1}$.

Bounds for $LNATE_{ap}$: From (3), $LNATE_{ap} = E[Y(1, S(0))|ap] - E[Y(0)|ap]$. Ass. B2 directly implies $LNATE_{ap} \geq 0$. Using the bounds previously obtained for the components in

[35] For brevity, in what follows we omit explicitly specifying when some quantities can be greater or lower than others unless we believe it is necessary. Hence, when min (or max) operators are present, it implies that none of the terms inside them is always lower (greater) than the other(s).

[36] The following equalities are helpful for the rest of the proofs. For scalars a, b, c and d we have: (i) $a - \max\{c, d\} = \min\{a-c, a-d\}$; (ii) $a - \min\{c, d\} = \max\{a-c, a-d\}$; (iii) $\max\{a, b\} - c = \max\{a-c, b-c\}$; (iv) $\min\{a, b\} - c = \min\{a-c, b-c\}$; (v) $\max\{a, b\} - \min\{c, d\} = \max\{a-c, a-d, b-c, b-d\}$; (vi) $\min\{a, b\} - \max\{c, d\} = \min\{a-c, a-d, b-c, b-d\}$.

$LNATE_{ap}$ we obtain two additional lower bounds: $L^{0,ap}-U^{0,ap}$ and $L^{0,ap}-U^{1,ap}$. By definition, $L^{0,ap}-U^{0,ap}\leq 0$. Also, employing Ass. B we have $U^{1,ap}\geq E[Y(1)|ap]\geq E[Y(0)|ap]\geq L^{0,ap}$, so $L^{0,ap}-U^{1,ap}\leq 0$. Hence, the lower bound for $LNATE_{ap}$ is 0. Using the bounds previously derived for the components of $LNATE_{ap}$, we have the upper bound is $U^{1,ap}-L^{0,ap}$. Thus, $0\leq LNATE_{ap}\leq (U^{1,ap}-L^{0,ap})$.

Bounds for $LMATE_{ap}$: $LMATE_{ap}=E[Y(1)|ap]-E[Y(1,S(0))|ap]$. Ass. B1 directly implies $LMATE_{ap}\geq 0$. Using the bounds previously obtained for the components of $LMATE_{ap}$ we obtain two additional lower bounds: $L^{1,ap}-U^{1,ap}$ and $L^{0,ap}-U^{1,ap}$. Since $L^{1,ap}-U^{1,ap}\leq 0$ (by definition) and $L^{0,ap}-U^{1,ap}\leq 0$ (from above), the lower bound for $LMATE_{ap}$ is 0. Using the bounds previously derived for the components of $LMATE_{ap}$, we have the upper bound is $U^{1,ap}-L^{0,ap}$. Thus, $0\leq LMATE_{ap}\leq (U^{1,ap}-L^{0,ap})$.

We now derive the bounds for $NATE$, starting with the upper bound. We use equations (6) to (9) to derive potential upper bounds for $NATE$ by plugging in the appropriate bounds derived above into the terms that are not point identified. The corresponding four potential upper bounds are:

$$\begin{aligned}
\Delta_1 &= E[Y|T=1]+\left(p_{1|1}-p_{1|0}\right)\left[U^{1,ap}-L^{0,ap}\right]-p_{0|1}L^{0,n0} \\
&\quad -p_{1|0}E[Y|T=0,S=1]-\left(p_{1|1}-p_{1|0}\right)\max\left\{L^{0,ap},L^{1,ap}\right\} \\
\Delta_2 &= p_{1|0}U^{1,n1}+p_{0|1}E[Y|T=1,S=0]+\left(p_{1|1}-p_{1|0}\right)U^{1,ap}-E[Y|T=0] \\
\Delta_3 &= E[Y|T=1]-E[Y|T=0] \\
\Delta_4 &= p_{1|0}U^{n1}+p_{0|1}U^{n0}+\left(p_{1|1}-p_{1|0}\right)\left[U^{1,ap}-L^{0,ap}\right]
\end{aligned}$$

After some algebra, we have $\Delta_1-\Delta_3=\pi_{ap}(U^{1,ap}-\max\left\{L^{0,ap},L^{1,ap}\right\})+\pi_{n0}(E[Y(0)|n0]-L^{0,n0})+\pi_{ap}(E[Y(0)|ap]-L^{0,ap})$; $\Delta_2-\Delta_3=\pi_{n1}(U^{1,n1}-E[Y(1)|n1])+\pi_{ap}(U^{1,ap}-E[Y(1)|ap])$ and $\Delta_4-\Delta_3=(\Delta_2-\Delta_3)+\pi_{n0}(E[Y(0)|n0]-L^{0,n0})+\pi_{ap}(E[Y(0)|ap]-L^{0,ap})$. Using the inequalities in Prop. 1 and the fact that $U^{1,ap}\geq\max\left\{L^{0,ap},L^{1,ap}\right\}$ (see above) we have: $\Delta_1-\Delta_3\geq 0$, $\Delta_2-\Delta_3\geq 0$ and $\Delta_4-\Delta_3\geq 0$. Hence, the upper bound for $NATE$ is $\Delta_3=E[Y|T=1]-E[Y|T=0]$.

Now consider the lower bound for $NATE$. Plugging in the bounds derived above for the corresponding non-point identified terms in equations (6) to (9) yields the lower bounds L^1, L^2, L^3 and L^4, as given in Prop. 2. After some algebra, we can write: $L^1-L^4=p_{1|0}(U^{1,ap}-E[Y|T=0,S=1]-\max\{0,L^{1,n1}-E[Y|T=0,S=1]\})-p_{1|1}(U^{1,ap}-E[Y|T=1,S=1])$; $L^2-L^4=p_{0|1}(E[Y|T=1,S=0]-L^{0,ap}-\max\{0,E[Y|T=1,S=0]-U^{0,n0}\})-p_{0|0}(E[Y|T=0,S=0]-L^{0,ap})$; $L^3-L^4=(L^1-L^4)+(L^2-L^4)$; $L^3-L^1=(L^2-L^4)$; $L^3-L^2=(L^1-L^4)$; and $L^1-L^2=(L^1-L^4)+(L^4-L^2)$. All six comparisons can be greater or less than zero depending on the data, so no potential lower bound is dropped. To show this, it is enough to get for each comparison one case where the difference can be greater or less than zero. For instance, consider the first difference. L^1-L^4 can be greater or less than zero if $\max\{0,L^{1,n1}-E[Y|T=$

$0, S = 1]\} = 0$ and $E[Y|T = 1, S = 1] \geq E[Y|T = 0, S = 1]$, since $p_{1|1} \geq p_{1|0} \geq 0$ and $(U^{1,ap} - E[Y|T = 0, S = 1]) \geq (U^{1,ap} - E[Y|T = 1, S = 1]) \geq 0$. Similar arguments can be made for the rest of the comparisons.

Finally, the bounds for $MATE$ follow directly from the bounds for $NATE$ and the fact that $MATE = ATE - NATE$. Q.D.E.

Proof of Proposition 3. As before, we first derive bounds for the non-point identified mean potential outcomes of the stratas, and for all the local net and mechanism average treatment effects.

Bounds for $E[Y(0)|n0]$: As discussed in the text in the paragraph before Proposition 3, Ass. C3 and equation (4) imply $E[Y(0)|n0] \leq E[Y|T = 0, S = 0]$. Since by definition $E[Y|T = 0, S = 0] \leq U^{0,n0}$, the upper bound in this case is $E[Y|T = 0, S = 0]$. Ass. C does not provide any additional information for a lower bound of $E[Y(0)|n0]$. Thus, $L^{0,n0} \leq E[Y(0)|n0] \leq E[Y|T = 0, S = 0]$.

Bounds for $E[Y(1)|n1]$: Ass. C1 and C2 imply $E[Y(1)|n1] \geq E[Y(1)|n0] = E[Y|T = 1, S = 0]$. Ass. C6 and equation (5) yield $E[Y(1)|n1] \geq E[Y|T = 1, S = 1]$. By definition, $E[Y|T = 1, S = 1] \geq L^{1,n1}$, and by (10) $E[Y|T = 1, S = 1] \geq E[Y|T = 1, S = 0]$. Since Ass. C does not provide any additional information for an upper bound of $E[Y(1)|n1]$, we have $E[Y|T = 1, S = 1] \leq E[Y(1)|n1] \leq U^{1,n1}$.

Bounds for $E[Y(0)|ap]$: Ass. C3 and equation (4) yield $E[Y(0)|ap] \geq E[Y|T = 0, S = 0]$, where by definition $E[Y|T = 0, S = 0] \geq L^{0,ap}$. As for the upper bound, Ass. C4 implies $E[Y(0)|ap] \leq E[Y(0)|n1] = E[Y|T = 0, S = 1]$, which can be greater or less than $U^{0,ap}$. Thus, $E[Y|T = 0, S = 0] \leq E[Y(0)|ap] \leq \min\{U^{0,ap}, E[Y|T = 0, S = 1]\}$.

Bounds for $E[Y(1)|ap]$: Ass. C6 and equation (5) yield $E[Y(1)|ap] \leq E[Y|T = 1, S = 1]$, where by definition $U^{1,ap} \geq E[Y|T = 1, S = 1]$. As for the lower bound, Ass. C5 implies $E[Y(1)|ap] \geq E[Y(1)|n0] = E[Y|T = 1, S = 0]$, which can be greater or less than $L^{1,ap}$. Thus, $\max\{L^{1,ap}, E[Y|T = 1, S = 0]\} \leq E[Y(1)|ap] \leq E[Y|T = 1, S = 1]$.

Bounds for $E[Y(1, S(0))|ap]$: Ass. C1 implies $E[Y(1, S(0))|ap] \geq E[Y(1)|n0] = E[Y|T = 1, S = 0]$. Combining Ass. C2 with the bounds previously derived for $E[Y(1)|n1]$ yields $E[Y(1, S(0))|ap] \leq E[Y(1)|n1] \leq U^{1,n1}$. Hence, $E[Y|T = 1, S = 0] \leq E[Y(1, S(0))|ap] \leq U^{1,n1}$. The bounds for $LNATE_{n0}$, $LNATE_{n1}$, $LNATE_{ap}$ and $LMATE_{ap}$ follow directly by plugging in the appropriate bounds previously derived for each of their non-point identified components. For instance, for $LNATE_{n0} = E[Y|T = 1, S = 0] - E[Y(0)|n0]$ we employ the bounds previously derived for $E[Y(0)|n0]$ to get $(E[Y|T = 1, S = 0] - E[Y|T = 0, S = 0]) \leq LNATE_{n0} \leq U^{n0}$.

We now derive the bounds for $NATE$, starting with the upper bound. As before, we use equations (6) to (9) to derive potential upper bounds for $NATE$ by plugging in the appropriate bounds derived above into the terms that are not point identified. The corresponding four

potential upper bounds are:

$$\begin{aligned}
\Delta_1 &= E\left[Y|T=1\right] + \left(p_{1|1} - p_{1|0}\right)\overline{U}^{ap} - p_{0|1}L^{0,n0} - p_{1|0}E\left[Y|T=0,S=1\right] \\
&\quad - \left(p_{1|1} - p_{1|0}\right)\max\{L^{1,ap}, E[Y|T=1,S=0]\} \\
\Delta_2 &= p_{1|0}U^{1,n1} + p_{0|1}E\left[Y|T=1,S=0\right] + \left(p_{1|1} - p_{1|0}\right)U^{1,n1} - E\left[Y|T=0\right] \\
\Delta_3 &= E\left[Y|T=1\right] - E\left[Y|T=0\right] - \left(p_{1|1} - p_{1|0}\right)\overline{L}_m^{ap} \\
\Delta_4 &= p_{1|0}U^{n1} + p_{0|1}U^{n0} + \left(p_{1|1} - p_{1|0}\right)\overline{U}^{ap}
\end{aligned}$$

After some algebra we obtain $\Delta_3 - \Delta_1 = \Delta_2 - \Delta_4 = p_{0|1}(L^{0,n0} - E[Y|T=0,S=0]) \leq 0$, since by definition $L^{0,n0} \leq E[Y|T=0,S=0]$. We also obtain that $\Delta_3 - \Delta_2 = (p_{1|1} - p_{1|0})(U^{1,n1} - \max\{L^{1,ap}, E[Y|T=1,S=0]\}) - p_{1|1}(U^{1,n1} - E[Y|T=1,S=1])$. Note that: (i) $p_{1|1} \geq (p_{1|1} - p_{1|0}) = \pi_{ap} \geq 0$; (ii) $U^{1,n1} - E[Y|T=1,S=1] \geq 0$ by definition of $U^{1,n1}$; and $(U^{1,n1} - \max\{L^{1,ap}, E[Y|T=1,S=0]\}) \geq 0$ since $U^{1,n1} \geq E[Y|T=1,S=1] \geq L^{1,ap}$ (by definition) and $U^{1,n1} \geq E[Y|T=1,S=1] \geq E[Y|T=1,S=0]$ by (10); (iii) $(U^{1,n1} - \max\{L^{1,ap}, E[Y|T=1,S=0]\}) \geq (U^{1,n1} - E[Y|T=1,S=1])$, since $E[Y|T=1,S=1] \geq \max\{L^{1,ap}, E[Y|T=1,S=0]\}$ (see part ii). Parts (i) to (iii) imply that $\Delta_3 - \Delta_2$ can be greater or less than zero. Thus, the upper bound of $NATE$ is: $\min\{\overline{U}^1, \overline{U}^2\}$, where we let $\overline{U}^1 = \Delta_2$ and $\overline{U}^2 = \Delta_3$.

Now consider the lower bound for $NATE$. Plugging in the bounds derived above for the corresponding non-point identified terms in equations (6) to (9) yields the following potential lower bounds:

$$\begin{aligned}
\Upsilon_1 &= E\left[Y|T=1\right] + \left(p_{1|1} - p_{1|0}\right)\overline{L}^{ap} - p_{0|1}E\left[Y|T=0,S=0\right] \\
&\quad - p_{1|0}E\left[Y|T=0,S=1\right] - \left(p_{1|1} - p_{1|0}\right)E\left[Y|T=1,S=1\right] \\
\Upsilon_2 &= p_{1|0}E\left[Y|T=1,S=1\right] + p_{0|1}E\left[Y|T=1,S=0\right] \\
&\quad + \left(p_{1|1} - p_{1|0}\right)E\left[Y|T=1,S=0\right] - E\left[Y|T=0\right] \\
\Upsilon_3 &= E\left[Y|T=1\right] - E\left[Y|T=0\right] - \left(p_{1|1} - p_{1|0}\right)\overline{U}_m^{ap} \\
\Upsilon_4 &= p_{1|0}\overline{L}^{n1} + p_{0|1}\overline{L}^{n0} + \left(p_{1|1} - p_{1|0}\right)\overline{L}^{ap}
\end{aligned}$$

After some algebra we obtain $\Upsilon_1 = \Upsilon_4$, $\Upsilon_2 = \Upsilon_3$ and $\Upsilon_4 - \Upsilon_3 = \pi_{ap}(E[Y|T=0,S=0] - \min\{U^{0,ap}, E[Y|T=0,S=1]\}) \leq 0$, since by definition $U^{0,ap} \geq E[Y|T=0,S=0]$ and by (10) $E[Y|T=0,S=1] \geq E[Y|T=0,S=0]$. Thus, the lower bound for $NATE$ equals $\overline{L} \equiv \Upsilon_3$.

Finally, the bounds for $MATE$ follow directly from the bounds for $NATE$ and the fact that $MATE = ATE - NATE$. Q.D.E.

Proof of Proposition 4. *Bounds for* $E[Y(0)|n0]$: Ass. B2 implies $E[Y(0)|n0] \leq E[Y(1)|n0] = E[Y|T=1,S=0]$; and Ass. C3 implies $E[Y(0)|n0] \leq E[Y|T=0,S=0]$

(see proof of Prop. 3), where by definition $U^{0,n0} \geq E[Y|T=0,S=0]$. Combining the rest of the assumptions does not yield any additional upper bound for $E[Y(0)|n0]$ that could be lower than $E[Y|T=0,S=0]$ or $E[Y|T=1,S=0]$.[37] Equation (4) and the fact that $E[Y(1)|n0] = E[Y|T=1,S=0]$ imply that $E[Y|T=1,S=0]$ can be greater or less than $E[Y|T=0,S=0]$ since, even though $E[Y(0)|n0] \leq E[Y(1)|n0]$ (by Ass. B2), we have that $E[Y(1)|n0]$ can be greater or less than $E[Y(0)|ap]$. Hence, the upper bound for $E[Y(0)|n0]$ is $\min\{E[Y|T=1,S=0], E[Y|T=0,S=0]\}$. Ass. B and C do not provide any additional information for a lower bound of $E[Y(0)|n0]$. Thus, $L^{0,n0} \leq E[Y(0)|n0] \leq \min\{E[Y|T=1,S=0], E[Y|T=0,S=0]\}$.

Bounds for $E[Y(1)|n1]$: Ass. B2 implies $E[Y(1)|n1] \geq E[Y(0)|n1] = E[Y|T=0,S=1]$; and Ass. C6 implies $E[Y(1)|n1] \geq E[Y|T=1,S=1]$ (see proof of Prop. 3), where by definition $E[Y|T=1,S=1] \geq L^{1,n1}$. Combining Assumptions B and C does not yield any additional lower bound for $E[Y(1)|n1]$ that could be greater than $E[Y|T=1,S=1]$ or $E[Y|T=0,S=1]$. Equation (5) and the fact that $E[Y(0)|n1] = E[Y|T=0,S=1]$ imply that $E[Y|T=0,S=1]$ can be greater or less than $E[Y|T=1,S=1]$ since, even though $E[Y(0)|n1] \leq E[Y(1)|n1]$ (by Ass. B2), we have that $E[Y(0)|n1]$ can be greater or less than $E[Y(1)|ap]$. Hence, the lower bound for $E[Y(1)|n1]$ is $\max\{E[Y|T=1,S=1], E[Y|T=0,S=1]\}$. Ass. B and C do not provide any additional information for an upper bound of $E[Y(1)|n1]$. Thus, $\max\{E[Y|T=1,S=1], E[Y|T=0,S=1]\} \leq E[Y(1)|n1] \leq U^{1,n1}$.

Bounds for $E[Y(0)|ap]$: Ass. B does not provide any information for a lower bound of $E[Y(0)|ap]$; while Ass. C3 and equation (4) yield $E[Y(0)|ap] \geq E[Y|T=0,S=0]$, where by definition $E[Y|T=0,S=0] \geq L^{0,ap}$. Regarding an upper bound, Ass. A1 and A2 imply $E[Y(0)|ap] \leq U^{0,ap}$. Ass. C4 implies $E[Y(0)|ap] \leq E[Y(0)|n1] = E[Y|T=0,S=1]$. Finally, Ass. B implies $E[Y(1)|ap] \geq E[Y(0)|ap]$. Below we show that the upper bound for $E[Y(1)|ap]$ under Ass. A1, A2, B and C equals $E[Y|T=1,S=1]$, so $E[Y(0)|ap] \leq E[Y|T=1,S=1]$. Depending on the data, any of the previous three upper bounds for $E[Y(0)|ap]$ can be less than the other two. Thus, we obtain $E[Y|T=0,S=0] \leq E[Y(0)|ap] \leq \min\{U^{0,ap}, E[Y|T=0,S=1], E[Y|T=1,S=1]\}$.

Bounds for $E[Y(1)|ap]$: Ass. B does not provide any information for an upper bound of $E[Y(1)|ap]$; while Ass. C6 and equation (5) yield $E[Y(1)|ap] \leq E[Y|T=1,S=1]$, where by definition $E[Y|T=1,S=1] \leq U^{1,ap}$. Regarding a lower bound, Ass. A1 and A2 imply $E[Y(1)|ap] \geq L^{1,ap}$. Ass. C5 implies $E[Y(1)|ap] \geq E[Y(1)|n0] = E[Y|T=1,S=0]$. Finally, Ass. B implies $E[Y(1)|ap] \geq E[Y(0)|ap]$. Above we showed that the lower bound for $E[Y(0)|ap]$ under Ass. A1, A2, B and C equals $E[Y|T=0,S=0]$, so $E[Y(1)|ap] \geq E[Y|T=$

[37] For instance, combining Ass. C3, C4 and B2 yields $E[Y(1)|n1] \geq E[Y(0)|n1] \geq E[Y(0)|ap] \geq E[Y(0)|n0]$, which implies $E[Y(0)|n1] = E[Y|T=0,S=1] \geq E[Y(0)|n0]$ and $U^{1,n1} \geq E[Y(1)|n1] \geq E[Y(0)|n0]$. However, by (10) we have $E[Y|T=0,S=1] \geq E[Y|T=0,S=0]$; and by (11) we have: $U^{1,n1} \geq E[Y|T=1,S=1] \geq E[Y|T=0,S=0]$.

$0, S = 0]$. Depending on the data, any of the previous three lower bounds for $E[Y(1)|ap]$ can be greater than the other two. Thus, we obtain $\max\{L^{1,ap}, E[Y|T = 1, S = 0], E[Y|T = 0, S = 0]\} \leq E[Y(1)|ap] \leq E[Y|T = 1, S = 1]$.

Bounds for $E[Y(1, S(0))|ap]$: Ass. B2 implies $E[Y(1, S(0))|ap] \geq E[Y(0)|ap]$. From above, the lower bound for $E[Y(0)|ap]$ under Ass. A1, A2, B and C equals $E[Y|T = 0, S = 0]$. Ass. C1 implies $E[Y(1, S(0))|ap] \geq E[Y(1)|n0] = E[Y|T = 1, S = 0]$, which can be greater or less than $E[Y|T = 0, S = 0]$ (see above). Hence, $E[Y(1, S(0))|ap] \geq \max\{E[Y|T = 0, S = 0], E[Y|T = 1, S = 0]\}$. Ass. B1 implies $E[Y(1)|ap] \geq E[Y(1, S(0))|ap]$. From above, the upper bound for $E[Y(1)|ap]$ under Ass. A1, A2, B and C equals $E[Y|T = 1, S = 1]$. Note that C2 implies $E[Y(1, S(0))|ap] \leq E[Y(1)|n1] \leq U^{1,n1}$, but by definition $E[Y|T = 1, S = 1] \leq U^{1,n1}$. Therefore, $\max\{E[Y|T = 1, S = 0], E[Y|T = 0, S = 0]\} \leq E[Y(1, S(0))|ap] \leq E[Y|T = 1, S = 1]$.

Bounds for $LNATE_{n0}$: From (3), $LNATE_{n0} = E[Y|T = 1, S = 0] - E[Y(0)|n0]$. Using the bounds previously derived for $E[Y(0)|n0]$ we have: $\max\{0, \overline{L}^{n0}\} \leq LNATE_{n0} \leq U^{n0}$.

Bounds for $LNATE_{n1}$: From (3), $LNATE_{n1} = E[Y(1)|n1] - E[Y|T = 0, S = 1]$. Using the bounds previously derived for $E[Y(1)|n1]$ we have: $\max\{0, \overline{L}^{n1}\} \leq LNATE_{n1} \leq U^{n1}$.

Bounds for $LNATE_{ap}$: From (3), $LNATE_{ap} = E[Y(1, S(0))|ap] - E[Y(0)|ap]$. Ass. B2 directly implies $LNATE_{ap} \geq 0$. Using the bounds previously obtained for the components of $LNATE_{ap}$ we obtain six additional potential lower bounds: $E[Y|T = 1, S = 0] - U^{0,ap}$, $E[Y|T = 1, S = 0] - E[Y|T = 1, S = 1]$, $E[Y|T = 1, S = 0] - E[Y|T = 0, S = 1]$, $E[Y|T = 0, S = 0] - U^{0,ap}$, $E[Y|T = 0, S = 0] - E[Y|T = 1, S = 1]$ and $E[Y|T = 0, S = 0] - E[Y|T = 0, S = 1]$. Note that: $E[Y|T = 1, S = 0] - E[Y|T = 1, S = 1] \leq 0$ by (10); $E[Y|T = 0, S = 0] - U^{0,ap} \leq 0$ by definition; $E[Y|T = 0, S = 0] - E[Y|T = 1, S = 1] \leq 0$ by (11); and $E[Y|T = 0, S = 0] - E[Y|T = 0, S = 1] \leq 0$ by (10). Hence, $LNATE_{ap} \geq \max\{E[Y|T = 1, S = 0] - U^{0,ap}, E[Y|T = 1, S = 0] - E[Y|T = 0, S = 1], 0\} = \max\{0, \overline{L}^{ap}\}$. Using the bounds previously derived for the components of $LNATE_{ap}$, we have that the upper bound is $\widetilde{U}^{ap} = E[Y|T = 1, S = 1] - E[Y|T = 0, S = 0]$.

Bounds for $LMATE_{ap}$: $LMATE_{ap} = E[Y(1)|ap] - E[Y(1, S(0))|ap]$. Ass. B1 directly implies $LMATE_{ap} \geq 0$. Using the bounds previously obtained for the components of $LMATE_{ap}$ we obtain three additional potential lower bounds: $L^{1,ap} - E[Y|T = 1, S = 1]$, $E[Y|T = 1, S = 0] - E[Y|T = 1, S = 1]$ and $E[Y|T = 0, S = 0] - E[Y|T = 1, S = 1]$. Each of these three expressions is less than or equal to zero because of the definition of $L^{1,ap}$, (10) and (11), respectively. Using the bounds previously derived for the components of $LMATE_{ap}$ we have the upper bound is $E[Y|T = 1, S = 1] - \max\{E[Y|T = 1, S = 0], E[Y|T = 0, S = 0]\}$. Thus, $0 \leq LMATE_{ap} \leq \widetilde{U}_m^{ap}$.

We now derive the bounds for $NATE$, starting with the upper bound. As before, we use equations (6) to (9) and the bounds obtained above to derive potential upper bounds for $NATE$.

The corresponding four potential upper bounds are:

$$\begin{aligned}
\Delta_1 &= E\left[Y|T=1\right] + \left(p_{1|1} - p_{1|0}\right)\widetilde{U}^{ap} - p_{0|1}L^{0,n0} - p_{1|0}E\left[Y|T=0,S=1\right] \\
&\quad - \left(p_{1|1} - p_{1|0}\right)\max\{L^{1,ap}, E[Y|T=1,S=0], E[Y|T=0,S=0]\} \\
\Delta_2 &= p_{1|0}U^{1,n1} + p_{0|1}E\left[Y|T=1,S=0\right] + \left(p_{1|1} - p_{1|0}\right)E[Y|T=1,S=1] \\
&\quad - E\left[Y|T=0\right] \\
\Delta_3 &= E\left[Y|T=1\right] - E\left[Y|T=0\right] \\
\Delta_4 &= p_{1|0}U^{n1} + p_{0|1}U^{n0} + \left(p_{1|1} - p_{1|0}\right)\widetilde{U}^{ap}
\end{aligned}$$

After some algebra we obtain $\Delta_1 - \Delta_3 = \pi_{ap}(E[Y|T=1,S=1] - \max\{L^{1,ap}, E[Y|T=1,S=0], E[Y|T=0,S=0]\}) + p_{0|1}(E[Y|T=0,S=0] - L^{0,n0})$. By definition, $E[Y|T=0,S=0] \geq L^{0,n0}$ and $E[Y|T=1,S=1] \geq L^{1,ap}$. Also, $E[Y|T=1,S=1] \geq E[Y|T=1,S=0]$ and $E[Y|T=1,S=1] \geq E[Y|T=0,S=0]$ by (10) and (11), respectively. Hence, $\Delta_1 - \Delta_3 \geq 0$. We also have: $\Delta_2 - \Delta_3 = p_{1|0}(U^{1,n1} - E[Y|T=1,S=1]) \geq 0$, by definition of $U^{1,n1}$. Finally, we have $\Delta_4 - \Delta_3 = p_{1|0}(U^{1,n1} - E[Y|T=1,S=1]) + p_{0|1}(E[Y|T=0,S=0] - L^{0,n0}) \geq 0$. Thus, the upper bound for $NATE$ equals $\widetilde{U} \equiv \Delta_3 = E\left[Y|T=1\right] - E\left[Y|T=0\right]$.

Now consider the lower bound for $NATE$. Plugging in the bounds derived above for the corresponding non-point identified terms in equations (6) to (9) yields the following potential lower bounds:

$$\begin{aligned}
\Upsilon_1 &= E\left[Y|T=1\right] + \left(p_{1|1} - p_{1|0}\right)\max\{0, \overline{L}^{ap}\} \\
&\quad - p_{0|1}\min\{E\left[Y|T=0,S=0\right], E\left[Y|T=1,S=0\right]\} \\
&\quad - p_{1|0}E\left[Y|T=0,S=1\right] - \left(p_{1|1} - p_{1|0}\right)E[Y|T=1,S=1] \\
\Upsilon_2 &= p_{1|0}\max\{E\left[Y|T=1,S=1\right], E\left[Y|T=0,S=1\right]\} \\
&\quad + \left(p_{1|1} - p_{1|0}\right)\max\{E[Y|T=1,S=0], E[Y|T=0,S=0]\} \\
&\quad + p_{0|1}E\left[Y|T=1,S=0\right] - E\left[Y|T=0\right] \\
\Upsilon_3 &= E\left[Y|T=1\right] - E\left[Y|T=0\right] - \left(p_{1|1} - p_{1|0}\right)\widetilde{U}^{ap}_m \\
\Upsilon_4 &= p_{1|0}\max\{0, \overline{L}^{n1}\} + p_{0|1}\max\{0, \overline{L}^{n0}\} + \left(p_{1|1} - p_{1|0}\right)\max\{0, \overline{L}^{ap}\}
\end{aligned}$$

After some algebra we obtain $\Upsilon_1 - \Upsilon_4 = \Upsilon_3 - \Upsilon_2 = p_{1|0}\min\{\overline{L}^{n1}, 0\} \leq 0$. Υ_2 can be greater or less than Υ_4 depending on the data. As before, it is enough to show one case in which $\Upsilon_2 - \Upsilon_4$ is greater than zero and one in which it is less than zero. After some algebra we can write $\Upsilon_2 - \Upsilon_4 = p_{0|1}E\left[Y|T=1,S=0\right] + \pi_{ap}\max\{E[Y|T=1,S=0], E[Y|T=0,S=0]\} - p_{0|0}E[Y|T=0,S=0] - p_{0|1}\max\{0, \overline{L}^{n0}\} - \pi_{ap}\max\{0, \overline{L}^{ap}\}$. Let $\overline{L}^{n0} = E[Y|T=1,S=0] - E[Y|T=0,S=0] \leq 0$. Then, $\Upsilon_2 - \Upsilon_4 = p_{0|1}(E\left[Y|T=1,S=0\right] - E[Y|T=0,S=0]) - \pi_{ap}\max\{0, \overline{L}^{ap}\} \leq 0$. Now let $\overline{L}^{n0} = E[Y|T=1,S=0] - E[Y|T=0,S=0] \geq 0$. Then,

$\Upsilon_2 - \Upsilon_4 = \pi_{ap}(\overline{L}^{n0} - \max\{0, \overline{L}^{ap}\})$, which is greater or equal to zero if $\overline{L}^{ap} \leq 0$.[38] Thus, the lower bound for $NATE$ equals $\max\{\widetilde{L}^1, \widetilde{L}^2\}$, where $\widetilde{L}^1 \equiv \Upsilon_2$ and $\widetilde{L}^2 \equiv \Upsilon_4$.

Finally, the bounds for $MATE$ follow directly from the bounds for $NATE$ and the fact that $MATE = ATE - NATE$. Q.D.E.

References

[1] Andrews, D. and Soares, G. (2010), "Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection", *Econometrica*, 78, 119-57.

[2] Angrist, J., and Chen, S. (2008), "Long-Term Economic Consequences of Vietnam-Era Conscription: Schooling, Experience and Earnings", IZA Discussion Paper No. 3628.

[3] Angrist, J., Imbens, G., and Rubin, D. (1996), "Identification of Causal Effects Using Instrumental Variables", *Journal of the American Statistical Association*, 91, 444–472.

[4] Angrist, J., and Pischke, J-S. (2010), "The Credibility Revolution in Empirical Economics: How Better Research Design is Taking the Con out of Econometrics", *Journal of Economic Perspectives,* 24 (2), 3-30.

[5] Balke, A., and Pearl, J. (1997), "Bounds on Treatment Effects from Studies with Imperfect Compliance", *Journal of the American Statistical Association*, 92, 1171–1176.

[6] Black, D. and Smith, J. (2004), "How Robust is the Evidence on the Effects of College Quality? Evidence from Matching", *Journal of Econometrics*, 121, 99-124.

[7] Bugni, F. (2010), "Bootstrap Inference in Partially Identified Models Defined by Moment Inequalities: Coverage of the Identified Set", *Econometrica,* 78, 735-53.

[8] Cai, Z., Kuroki, M., Pearl, J. and Tian, J. (2008), "Bounds on Direct Effects in the Presence of Confounded Intermediate Variables", *Biometrics*, 64, 695-701.

[9] Chernozhukov, V., Hong, H. and Tamer, E. (2007), "Estimation and Confidence Regions for Parameter Sets in Econometric Models", *Econometrica*, 75, 1243-1284.

[10] Clark, C. (1961), "The Greatest of a Finite Set of Random Variable", *Operations Research*, 145-162.

[11] Currie, J. and Moretti, E. (2003), "Mother's Education and the Intergenerational Transmission of Human Capital: Evidence from College Openings", *Quarterly Journal of Economics*, 118(4), 1495-1532.

[38] In fact, it can be shown that $\Upsilon_2 - \Upsilon_4 = \pi_{ap}(\overline{L}^{n0} - \max\{0, \overline{L}^{ap}\}) \geq 0$ regardless of the value of $\overline{L}^{ap}$ as long as $\overline{L}^{n0} \geq 0$.

[12] Dearden, L. Ferri, J. and Meguir, C. (2002), "The Effect of School Quality on Educational Attainment and Wages" *Review of Economics and Statistics*, 84, 1-20.

[13] Deaton, A. (2010a), "Instruments, Randomization, and Learning about Development", *Journal of Economic Literature,* 48, 424-455.

[14] Deaton, A. (2010b), "Understanding the Mechanisms of Economic Development", NBER Working Paper 15891.

[15] Duflo, E., Glennerster, R., and Kremer, M. (2008), "Using Randomization in Development Economics Research: a Toolkit", in T.P. Schultz and J. Strauss (eds.) *Handbook of Development Economics*, Vol. 4, Elsevier Science North Holland, 3895–962.

[16] Flores, C. A. and Flores-Lagunes, A. (2009), "Identification and Estimation of Causal Mechanisms and Net Effects of a Treatment under Unconfoundedness", IZA Discussion paper No. 4237.

[17] Flores, C. A. and Flores-Lagunes, A. (2010), "Testing Implications of the Exclusion Restriction Assumption in Just-Identified Instrumental Variable Models", mimeo, Department of Economics, University of Miami.

[18] Flores-Lagunes, A., Gonzalez, A., and Neumann, T. (2010), "Learning but not Earning? The Impact of Job Corps Training on Hispanic Youth", *Economic Inquiry,* 48, 651-67.

[19] Frangakis,C.E. and Rubin D. (2002) "Principal Stratification in Causal Inference", *Biometrics*, 58, 21-29.

[20] Gallop, R., Small, D., Lin, J., Elliot, M., Joffe, M. and Ten Have, T. (2009), "Mediation Analysis with Principal Stratification", *Statistics in Medicine*, 28, 1108-1130.

[21] Heckman, J. (2010), "Building Bridges between Structural and Program Evaluation Approaches to Evaluating Policy", *Journal of Economic Literature,* 48, 356-398.

[22] Heckman, J., Smith, J. and Clements, N. (1997), "Making the Most Out of Programme Evaluations and Social Experiments: Accounting for Heterogeneity in Programme Impacts", *Review of Economic Studies*, 64(3), 487-535.

[23] Heckman, J., LaLonde, R. and Smith, J. (1999) "The Economics and Econometrics of Active Labor Market Programs" in O. Ashenfelter and D. Card (eds.) *Handbook of Labor Economics.* Elsevier Science North Holland, 1865-2097.

[24] Heckman, J. and Urzua, S. (2010), "Comparing IV with Structural Models: What Simple IV can and cannot Identify", *Journal of Econometrics*, 156, 27-37.

[25] Holland, P. (1986) "Statistics and Causal Inference", *Journal of the American Statistical Association*, 81, 945-70.

[26] Imai, K., Keele, L., and Yamamoto, T. (2010), "Identification, Inference, and Sensitivity Analysis for Causal Mediation Effects", Working Paper, Department of Politics, Princeton University.

[27] Imbens, G. (2010) "Better LATE Than Nothing: Some Comments on Deaton (2009) and Heckman and Urzua (2009)", *Journal of Economic Literature,* 48, 399-423.

[28] Imbens, G. W. and Angrist, J. D. (1994), "Identification and Estimation of Local Average Treatment Effects", *Econometrica*, 62 (2), 467-475.

[29] Imbens, G. and Wooldridge, J. (2009), "Recent Developments in the Econometrics of Program Evaluation", *Journal of Economic Literature*, 47, 5-86.

[30] Joffe, M., Small, D. and Hsu, C.-Y. (2007), "Defining and Estimating Intervention Effects for Groups that will Develop an Auxiliary Outcome", *Statistical Science*, 22, 74-97.

[31] Kaufman, S., Kaufman, J., MacLenose, R., Greenland, S., and Poole, C. (2005), "Improved Estimation of Controlled Direct Effects in the Presence of Unmeasured Confounding of Intermediate Variables", *Statistics in Medicine* 24, 1683-1702. (Correction, 25, 3228).

[32] Keane, M. (2010), "Structural vs. Atheoretic Approaches to Econometrics", *Journal of Econometrics*, 156, 3-20.

[33] Lee, D. (2009), "Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects", *Review of Economic Studies*, 76, 1071-102.

[34] Manski, C. (1997), "Monotone Treatment Response", *Econometrica*, 65, 1311-1334.

[35] Manski, C. (2003), *Partial Identification of Probability Distributions*, Springer Series in Statistics.

[36] Manski, C. (2007), *Identification for Prediction and Decision*, Harvard University Press.

[37] Manski, C. and Pepper, J. (2000), "Monotone Instrumental Variables: With an Application to the Returns to Schooling", *Econometrica*, 68, 997-1010.

[38] Mealli, F. and Rubin, D. (2003), "Assumptions Allowing the Estimation of Direct Causal Effects", *Journal of Econometrics*, 112, 79-87.

[39] Pearl, J. (2001), "Direct and Indirect Effects", *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence*, San Francisco, CA: Morgan Kaufmann, 411-20.

[40] Petersen, M., Sinisi, S., and van der Laan, M. (2006), "Estimation of Direct Causal Effects", *Epidemiology*, 17, 276-284.

[41] Robins, J. M. (2003), "Semantics of Causal DAG Models and the Identification of Direct and Indirect Effects", in *Highly Structured Stochastic Systems* (eds., P.J. Green, N.L. Hjort, and S. Richardson), 70-81. Oxford University Press, Oxford.

[42] Robins, J. and Greenland, S. (1992), "Identifiability and Exchangeability for Direct and Indirect Effects", *Epidemiology*, 3, 143-155.

[43] Robins J.M., Rotnitzky A. and Vansteelandt, S. (2007), Discussion of "Principal Stratification Designs to Estimate Input Data Missing Due to Death" by Frangakis, C., Rubin, D., An, M. and MacKenzie, E., *Biometrics*, 63, 650–654.

[44] Romano, J. and Shaikh, A. (2010), "Inference for the Identified Set in Partially Identified Econometric Models", *Econometrica*, 78 (1), 169-211.

[45] Rubin, D. (1980), Discussion of "Randomization Analysis of Experimental Data in the Fisher Randomization Test" by Basu, *Journal of the American Statistical Association*, 75, 591-93.

[46] Rubin, D. (2004), "Direct and Indirect Causal Effects via Potential Outcomes", *Scandinavian Journal of Statistics*, 31, 161-70.

[47] Rubin, D. (2005), "Causal Inference Using Potential Outcomes: Design, Modeling, Decisions", *Journal of the American Statistical Association*, 100, 322-331.

[48] Schochet, P., Burghardt, J., and Glazerman, S. (2001), *National Job Corps Study: The Impacts of Job Corps on Participants' Employment and Related Outcomes*. Princeton, NJ: Mathematica Policy Research, Inc.

[49] Simonsen, M. and Skipper, L. (2006), "The Costs of Motherhood: An Analysis Using Matching Estimators", *Journal of Applied Econometrics*, 21, 919-34.

[50] Sjölander, A. (2009), "Bounds on Natural Direct Effects in the Presence of Confounded Intermediate Variables", *Statistics in Medicine*, 28, 558-71.

[51] VanderWeele, T. J. (2008), "Simple Relations between Principal Stratification and Direct and Indirect Effects", *Statistics and Probability Letters*, 78, 2957-2962.

[52] Zhang, J.L. and Rubin, D. (2003), "Estimation of Causal Effects via Principal Stratification When Some Outcomes are Truncated by 'Death'", *Journal of Educational and Behavioral Statistics*, 28, 353-68.

[53] Zhang, J.L., Rubin, D. and Mealli, F. (2008), "Evaluating the Effects of Job Training Programs on Wages Through Principal Stratification", in D. Millimet et al. (eds) *Advances in Econometrics vol XXI*, Elsevier.

Table 2. Basic Point Estimates

Parameters	*Estimate*	*Standard Error*
(Total) Average Treatment Effects		
ATE program on employment	0.04	*(0.011)*
ATE program on earnings	18.11	*(4.759)*
ATE program on obtainment of degree	0.21	*(0.011)*
Strata Proportions		
π_{n0}	0.34	*(0.007)*
π_{n1}	0.45	*(0.009)*
π_{ap}	0.21	*(0.011)*
Conditional Probabilities		
Pr(S=0|T=0)	0.55	*(0.009)*
Pr(S=1|T=0)	0.45	*(0.009)*
Pr(S=0|T=1)	0.34	*(0.007)*
Pr(S=1|T=1)	0.66	*(0.007)*

	Employment		**Earnings**	
Conditional Means	*Estimate*	*Std. Error*	*Estimate*	*Std. Error*
E[Y|T=0, S=0]	0.57	*(0.012)*	149.05	*(4.323)*
E[Y|T=0, S=1]	0.66	*(0.013)*	197.93	*(6.092)*
E[Y|T=1, S=0]	0.55	*(0.012)*	142.78	*(4.548)*
E[Y|T=1, S=1]	0.70	*(0.008)*	213.28	*(3.729)*
Testable Implications				
E[Y|T=0, S=1]-E[Y|T=0, S=0]	0.09	*(0.018)*	48.87	*(7.365)*
E[Y|T=1, S=1]-E[Y|T=1, S=0]	0.15	*(0.014)*	70.50	*(5.894)*
E[Y|T=1, S=1]-E[Y|T=0, S=0]	0.13	*(0.015)*	64.23	*(5.708)*

Notes: Treatment is the random assignment to Job Corps; mechanism is the obtainment of a high school, GED, or vocational degree; outcomes are the weekly earnings and employment status in quarter 12 after randomization. Sample size is 8,020 with 2,975 control and 5,045 treatment individuals. Standard errors are based on 5,000 bootstrap replications.

Table 3. Estimated Bounds for the Employment Outcome

	Lower (LB) and Upper Bounds (UB) under Different Assumptions							
	Proposition 1 A1 and A2		Proposition 2 A1, A2 and B		Proposition 3 A1, A2, and C		Proposition 4 A1, A2, B and C	
Main Parameters	LB	UB	LB	UB	LB	UB	LB	UB
$LNATE_{n0}$	-0.365	0.241	0.000	0.241	-0.019	0.241	0.000	0.241
	(0.033)	*(0.029)*	*(0.000)*	*(0.029)*	*(0.017)*	*(0.029)*	*(0.004)*	*(0.029)*
$LNATE_{n1}$	-0.095	0.341	0.000	0.341	0.044	0.341	0.044	0.341
	(0.020)	*(0.014)*	*(0.000)*	*(0.014)*	*(0.015)*	*(0.014)*	*(0.015)*	*(0.014)*
$LNATE_{ap}$	--	--	0.000	1.000	-0.107	0.429	0.000	0.132
			(0.000)	*(0.001)*	*(0.018)*	*(0.013)*	*(0.000)*	*(0.015)*
$LMATE_{ap}$	--	--	0.000	1.000	-0.448	0.151	0.000	0.132
			(0.000)	*(0.001)*	*(0.013)*	*(0.014)*	*(0.000)*	*(0.014)*
NATE	--	--	0.000	0.041	0.009	0.134	0.020	0.041
			(0.000)	*(0.011)*	*(0.012)*	*(0.013)*	*(0.007)*	*(0.011)*
MATE	--	--	0.000	0.041	-0.094	0.032	0.000	0.021
			(0.000)	*(0.011)*	*(0.006)*	*(0.003)*	*(0.000)*	*(0.007)*

Notes: Treatment is the random assignment to Job Corps; mechanism is the obtainment of a high school, GED, or vocational degree; outcome is employment status in quarter 12 after randomization. Sample size is 8,020 with 2,975 control and 5,045 treatment individuals. In parenthesis are standard errors computed as described in footnote 32 in the text.

Table 4. Estimated Bounds for the Earnings Outcome

	Lower (LB) and Upper Bounds (UB) under Different Assumptions							
	Proposition 1 A1 and A2		Proposition 2 A1, A2 and B		Proposition 3 A1, A2, and C		Proposition 4 A1, A2, B and C	
Main Parameters	LB	UB	LB	UB	LB	UB	LB	UB
$LNATE_{n0}$	-96.41	111.09	0	111.09	-6.28	111.09	0	111.09
	(10.36)	*(6.90)*	*(0.00)*	*(6.90)*	*(6.25)*	*(6.90)*	*(1.00)*	*(6.90)*
$LNATE_{n1}$	-98.76	114.74	0	114.74	15.35	114.74	15.35	114.74
	(8.14)	*(10.31)*	*(0.00)*	*(10.31)*	*(7.09)*	*(10.31)*	*(6.78)*	*(10.31)*
$LNATE_{ap}$	--	--	0	456.65	-55.15	163.62	0	64.23
			(0.00)	*(10.66)*	*(7.71)*	*(9.21)*	*(0.00)*	*(5.71)*
$LMATE_{ap}$	--	--	0	456.65	-169.89	70.50	0	64.23
			(0.00)	*(10.66)*	*(9.38)*	*(5.89)*	*(0.00)*	*(5.42)*
NATE	--	--	0	18.11	3.37	53.63	6.85	18.11
			(0.00)	*(4.76)*	*(4.80)*	*(6.41)*	*(3.30)*	*(4.76)*
MATE	--	--	0	18.11	-35.52	14.74	0	11.26
			(0.00)	*(4.76)*	*(3.51)*	*(1.46)*	*(0.00)*	*(2.68)*

Notes: Treatment is the random assignment to Job Corps; mechanism is the obtainment of a high school, GED, or vocational degree; outcome is weekly earnings in quarter 12 after randomization. Sample size is 8,020 with 2,975 control and 5,045 treatment individuals. In parenthesis are standard errors computed as described in footnote 32 in the text.